UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL

MICROBIAL STRUCTURE AND NITRITE REDUCING COMMUNITIES ACROSS THE ATLANTIC OCEAN

Miguel Fernandes Guerreiro

MESTRADO EM BIOINFORMÁTICA E BIOLOGIA COMPUTACIONAL
ESPECIALIDADE EM BIOLOGIA COMPUTACIONAL

Dissertação orientada por:
Gerhard Herndl e Octávio Paulo

2017

À memória de Leonel Fernandes (06/05/1930 - 14/02/2014),
o meu avô, que me introduziu à Matemática.

Acknowledgements

Para a conclusão desta tese, nada teria sido possivel sem o aval do meu Orientador externo, Dr. Gerhard Herndl, pela oportunidade de trabalhar no seu grupo de investigação, e de me ter recebido de braços abertos e depositado tanta confiança nas minhas capacidades e trabalho.

Ao programa ERASMUS$^+$, pelo apoio financeiro e em particular à Ana Pereira pela ajuda na informação necessária

À Dr^a. Eva Sintes, a minha supervisora, que me guiou no processo de desenvolvimento da tese.

Ao meu orientador interno, Dr. Octávio Paulo, pela ajuda e flexibilidade durante o ano curricular e na candidatura à bolsa de ERASMUS.

Aos meus colegas de trabalho, que me deram novos ângulos para os problemas que encontrava.

À minha família e amigos, pelo apoio emocional, e claro, pela afeição que nutrem por mim.

À Stella-Maria Schey, *for all the support and love and for making my life much better*

Abstract

The Atlantic Ocean is the second largest water mass on the globe, with a complex biogeochemical system. As such, prokaryots have available a myriad of different environments to exploit resources, namely, nitrogen compounds, in which they participate and manipulate to form the nitrogen cycle. In this study we aim to characterize the microbial phylogenetic and denitritification diversity throughout the Atlantic. For this, water samples were retrieved from four depths in 12 points across the Atlantic, to extract, sequence and analyse the *16S rRNA* and *nir*k genes using three different pipelines. In addition, environmental data was retrieved to attempt to predict the distribution of each microbe in the studied area, by developing a Mahalanobis distance based model. From the three pipelines, the most promising results came from Uparse. The results from *16S rRNA* showed that Archaea communites tended to segregate mostly on depth while Bacteria segregated in relationship to their regional distribution. While the results from *nir*k reveal that Archaeal *nir*K-a containing communities consisted of three main clusters and Archaeal *nir*K-b communities were composed of two clusters. Furthermore, Bacteria harboring *nir*K clustered mainly according to the oceanographic region. The developed model could not show similar results observed in the field, and could not be applied to the *nir*K dataset. Certain phila distribution were cartographed in this transect for the first time, and patterns of distribution visited in other studies could be observed (bipolar and lower latitude bound phila). Both Archaea *nir*K harbouring communities were depth-stratified which could be associated to the higher nutrient supply rates in the epi- and upper mesopelagic as compared to bathypelagic waters. In contrast, bacterial nirK harbouring communities clustered according to the oceanographic regions probably

due to water mass formations. In the future, further work must be employed in the development of distribution explaining models.

Keywords: Modelling, Atlantic Ocean, Nitrite Reduction, Prokariota, Biodiversity

Resumo

O Oceano Atlântico é o segundo maior oceano do nosso planeta, e consiste em duas bacias em forma de S, separadas longitudinalmente pela Crista Médio-Atlântica. Estas formações geológicas influenciam as correntes, as temperaturas e a composição química da água, e consequentemente, as formas de vida que a habitam. A Crista Médio-Atlântica é caracterizada, ao longo de toda a sua extensão, pela presença de atividade vulcânica, que introduz na coluna de água enormes quantidades de metais e minerais a altas temperaturas. De notar a existência de zonas de alta produtividade como o Mar dos Sargaços, e a presença da corrente termo-halina, onde a água afunda na proximidade dos polos e volta a emergir nos oceanos Indico e Pacífico. A árvore da vida divide-se em três domínios: Eucariota, Bactéria e Arqueias. Arqueia e Bacteria formam o grupo denominado Procariota, o qual ocupa todos os habitats presentes no oceano e, por isso, constitui a maior porção de biomassa e faz parte central de vários sistemas biogeoquímicos. O Azoto está presente nos seres vivos em vários componentes orgânicos. Transforma-se em diferentes formas químicas no processo conhecido como "Ciclo do Azoto". Estes processos são em grande parte mediados por procariotas e, de especial interesse para este estudo, encontra-se o processo de desnitrificação, no qual o Nitrato, numa série de passos de redução, se torna em Azoto atmosférico. Atualmente, os microbiologistas utilizam diversas ferramentas moleculares para estimar a diversidade e identificar os micróbios, com base nos quais podemos começar a fazer estudos de biogeografia microbiana, biodiversidade, etc. Apesar dos padrões já identificados para estes seres, os fatores por tràs dos padrões biogeográficos continuam pouco caracterizados, estando ainda em discussão se os que predominam são o nicho ecológico clàssico e competição ou se, de acordo

com a teoria neutral, serão a dispersão e a deriva. A diversidade filogenética não fornece informação acerca do funcionamento do ecossistema. Os genes funcionais podem ser estudados neste sentido, como é o exemplo do gene *nir*, responsável pela redução de nitrito (desnitrificação) dentro do ciclo de Azoto. Foram desenvolvidos modelos para explicar os mecanismos causadores de diversos fenómenos, os quais fornecem simulações sem serem necessárias futuras observações no alvo em estudo. O presente estudo tem como objetivo caracterizar a diversidade filogenética e as comunidades procariotas envolvidas no processo de desnitrificação, ao longo do Oceano Atlântico e verificar a aplicação de um modelo preditivo da distribuição das mesmas. Procedeu-se à colheita de amostras em duas expedições (Geotraces-1 e -2) ao longo de um transepto no sentido norte-sul no lado Ocidental do Oceano Atlântico. Em 51 estações entre as latitudes 65°N e 55°S foram colhidas 4 a 8 amostras em profundidades desde a superfície até 4000 metros de profundidade. Para a colheita das amostras de água foram utilizadas garrafas *Niskin* , acopladas a um sensor CTD. As diferentes regiões oceanográficas amostradas foram classificadas da seguinte forma:*North Atlantic Arctic Province (ARCT), North Atlantic Drift Province (NADR), North Atlantic Gyral Province (NAG), Western Tropical Atlantic (WTRA), South Atlantic Gyral (SATL) e Subantarctic Province (SANT).* Para caracterizar a comunidade procariota, foram utilizadas amostras de quatro profundidades diferentes, colhidas em duas estações em cada região oceanográfica, de forma a sequenciar *nir*K e *16S rRNA* de Bactérias e Arqueias com tecnologia *Illumina.* Após a extração, amplificação e sequenciação dos genes *nir*K e textit16S rRNA, foram utilizadas três *pipelines* (Uparse, Qiime e Mothur) para calcular a tabela de Unidades Taxonómicas Operacionais (*Operational Taxonomic Units* - OTUs) e os correspondentes índices de biodiversidade, para comparação. Os resultados destas *pipelines* foram semelhantes entre si, pelo que apenas os resultados vindos de *Uparse* foram utilizados no desenvolvimento posterior do estudo. Esta *pipeline* usa ferramentas e comandos de *usearch*, já que

Uparse faz parte desta coleção de *software*. Para atribuir taxonomias às diferentes unidades, foi feito um *blast* com as sequências representativas das OTUs para comparar com a base de dados SILVA_123 (para *16S rRNA*) e uma base de dados desenvolvida (para *nir*K) no departamento de Microbiologia Oceanográfica da Universidade de Viena. Para melhor identificar padrões de diversidade, foram calculados, para cada amostra, índices de diversidade e equitatividade de Shannon e o número de Chao. O modelo desenvolvido para prever a distribuição das OTUs, calculou para cada uma delas, a distribuição normal multidimensional das variáveis ambientais observadas nos locais em que a unidade em estudo estava presente, e ponderada pela quantidade de OTUs aí registadas. De seguida, a distância de Mahalanobis foi medida entre a média multidimensional e cada local amostrado. Esta distância foi usada como aproximação à variância, e usada para calcular a quantidade de OTUs em cada amostra, através da sua subtração à média quantidade da OTU. Da sequenciação das nossas amostras resultaram 4 111 873 sequências. As pipelines Uparse, Qiime e Mothur obtiveram respetivamente 1 986, 7 941 e 244 734 OTUs das sequências das bactérias e o tempo de processamento foi aproximadamente 3 horas, 2 dias e 2-3 semanas respetivamente, com os índices de biodiversidade semelhantes para os produtos das três *pipelines*. Os filos de Arquias presentes no Oceano Atlântico incluem Euryarchaeota, Woesearchaeota (DHVEG-6), Thaumarchaeota, Aigarchaeota e o grupo marinho de fontes hidrotermais (MHVG). O mais representado foi o Taumarchaeota (51,7%), seguido do Euryarchaeota (45,2%). Em termos de abundância, o filo Euryarchaeota aumenta relativamente ao resto da comunidade de Arquias, em baixas latitudes, no ambiente epipelágico, em contraste com Thaumarchaeota e Euryarchaeota que habitam o meio batipelágico inferior. As comunidades de Arquias agruparam-se de acordo com a batimetria em que foram encontradas, e encontrando-se mais próximas de comunidades de batimetrias semelhantes. Os filos de bactérias presentes no Oceano Atlântico incluem Proteobacteria, Cyanobacteria, Actinobacteria, Chlo-

roflexi, Bacteroidetes, Acidobacteria, Planctomycetes, Marinimicrobia (SAR406), Nitrospirae, Parcubacteria, Verrucomicrobia, LCP-89, Lentisphaerae, Hydrogenedentes, Deinococcus-Thermus, Gracilibacteria, Spirochaetes, PAUC34f, Gemmatimonadetes, Sacharibacteria, Deferribacteres e Firmicutes. O mais representado foi Proteobactéria (63,7%). O filo Chloroflexi aumentou em profundidade e em baixas latitudes. O Actinobacteria foi mais abundante à superfície, e com redução da abundância nas regiões SANT, WTRA e ARCT. À superfície, o filo Bacteroidetes apresentou maiores abundâncias no hemisfério norte, enquanto em profundidade, foi mais abundante em altas latitudes. O Cyanobacteria esteve presente nas camadas superficiais, em abundâncias relativamente altas (11,6%). Tal como as Arquias, as comunidades bacterianas também se agruparam de acordo com a profundidade. Contudo as regiões ARCT e WTRA formaram um grupo à parte. As Arqueias apresentam duas formas de *nir*K: tipo a e tipo b, enquanto as Bactérias apenas possuem um. As comunidades de Arqueias que possuem tipo "a" apresentaram três agrupamentos: comunidades do ARCT-meio batipelágico superior-SANT mesopelágico, meio batipelágico inferior e o terceiro com as comunidades presentes nas camadas superficiais. As do tipo "b" apresentaram dois grupos: um com as comunidades do meio batipelágico superior e a comunidade do ARCT do meio batipelágico inferior, e a outra com as comunidades do meio batipelágico inferior. As comunidades bacterianas que possuem *nir*K agruparam-se principalmente segunda as regiões, com comunidades ARCT e NAG bem definidas, e um terceiro grupo com as comunidades do meio batipelágico inferior. O modelo aqui desenvolvido resultou em índices de diversidade inferiores aos observados e, no caso dos dados das Arqueias, os mesmos padrões de profundidade nos índices Chao (superficie-batipelágico inferior menor que batipelágico superior) entre os dados observados e os do modelo. Os índices de Shannon para os dados do modelo das bactérias, apresentaram um aumento com a profundidade, o mesmo padrão que é observado para os índices de Chao dos dados originais recolhidos. Neste

estudo foi cartografada, pela primeira vez, a distribuição de vários filos ao longo do Oceano Atlântico, tendo sido também confirmados os padrões de distribuição em outros estudos (bipolar e tropicalismo). Ambas as comunidades de arqueias com *nir*K, estão estratificadas ao longo das camadas do Oceano, o que pode estar associado a maiores taxas de fornecimento de nutrientes à superfície em comparação com o meio batipelágico. Em contraste, comunidades de bactérias com *nir*K, diferenciaram-se de acordo com a região, provavelmente devido às formações dominantes de massas de água presentes nestas áreas, que não se misturam entre si, possibilitando segregação. O modelo mostrou-se adequado em algumas situações mas revelou debilidades noutros casos, necessitando de melhorias para permitir extrapolações de dados a nível global.

Palavras Chave: Modulação, Oceano Atlântico, Redução de Nitrito, Procariota, Biodiversidade

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Atlantic Ocean

The Atlantic Ocean is the second largest ocean, and is delimited by the continents of Africa, America and Europe and by the Southern and Arctic Ocean (Figure 1.1) . It consists of two S-shaped basins separated longitudinally by the Mid-Atlantic Ridge, with minor transverse ridges generating a series of basins. These geological features influence the current systems, temperature and chemistry of the water, and therefore the life forms inhabiting it. The Mid-Atlantic Ridge is a submarine mountain range, sometimes reaching the sea-surface where it forms islands and archipelagos, such as Iceland and the Azores archipelago. Numerous active volcanoes and hydrothermal vents are located along the oceanic ridges, mainly underwater, releasing copious amounts of minerals and metals to the water column at high temperatures. The continental shelves are characterized by high biological productivity, as the hydrodynamics of these regions favour photosynthetic organisms staying within the euphotic zone, the sunlit surface waters where net primary production is possible. A particular feature is the Sargasso Sea in the North Atlantic Gyre, characterized by the presence of *Sargassum* seaweed in the surface waters.

Surface currents are wind driven, and therefore, due to the Ekman [and Coriolis] effects, generate two major oceanic gyres north and south of the Equator, with clockwise and counter clockwise rotation, respectively. Downwelling processes occur in the centre of the gyres and in the subpolar regions, while upwelling prevails

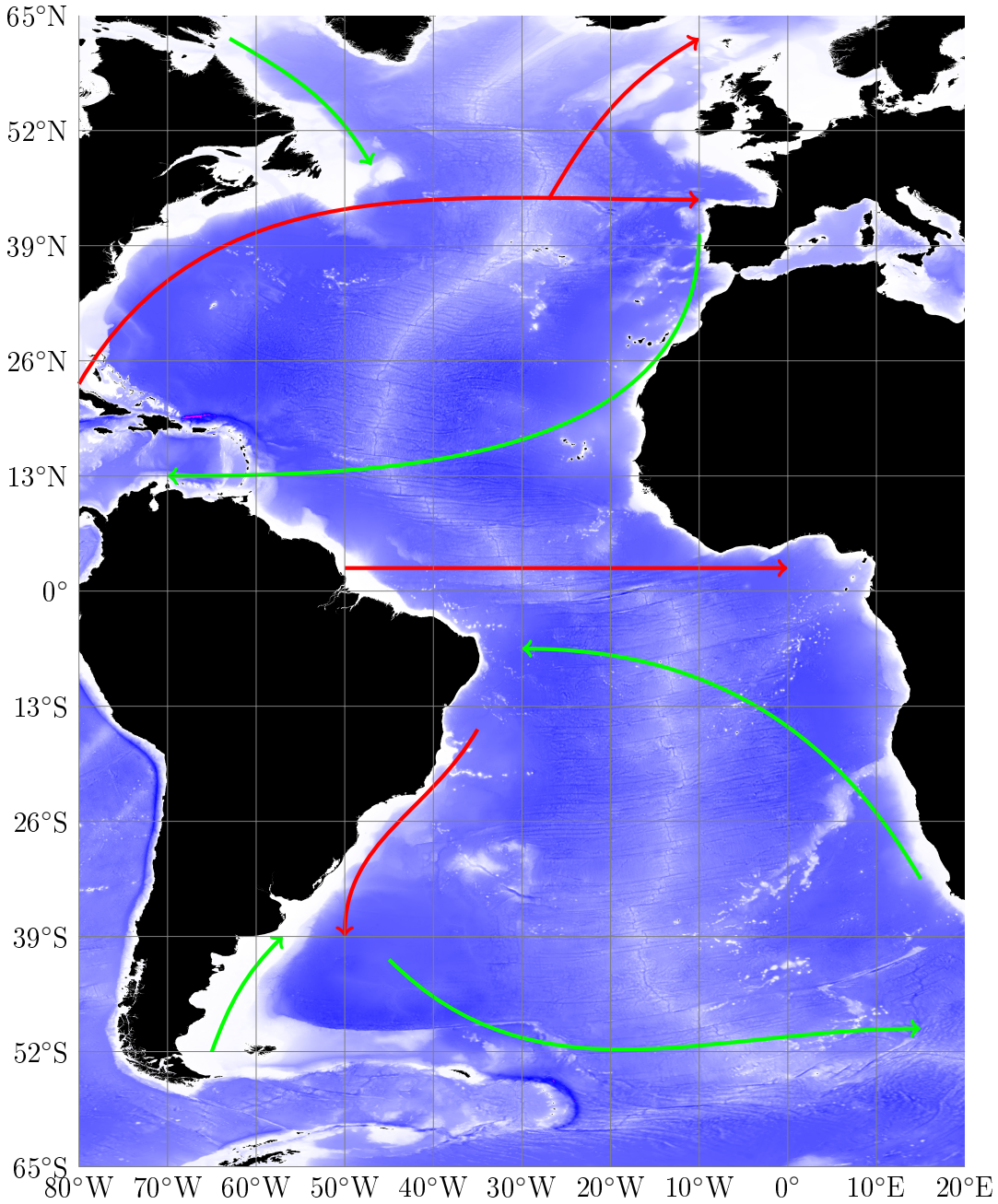

Figure 1.1: Map of the Atlantic Ocean. Land is colored in black, while ocean bathymetry is represented by a gradient from white to dark blue (from sea-surface level to 8000 meters bellow the surface). Depths bellow 8000 meters are colored in purple. Gross representation of surface currents are shown with arrows: → warmer currents than atmospheric temperature, → colder currents than atmospheric temperature. Bathymetric data retrieved from the General Bathymetric Chart of the Oceans website(`http://www.gebco.net/`, 2014)

at the equator. In contrast, the currents in the deep ocean are density driven, coined the thermohaline circulation (Rahmstorf, 2006). Sea surface water cools down on its way to the poles and becoming denser, eventually sinks at high latitudes, a process known as deep-water formation. The newly formed deep-water flows towards lower latitudes and is upwelled in the North Pacific and northern Indian Ocean.

1.2 Microbes

The tree of life has been traditionally separated into 3 domains of life (Figure 1.2) (Woese *et al.*, 1990). Eukaryotes, including humans, are formed by one or more cells which contain several compartments, such as the nucleus, where (most of) the genetic material of the organism is found. Bacteria are single-celled organisms without nucleus and a different cellular structure as compared to the eukaryotes, especially in the outer layer of the cell. The third domain of life, Archaea, consists of single-celled organisms without cellular compartments, however, they share with eukaryotes many molecular features (Woese *et al.*, 1990).

Archaea and Bacteria, frequently referred to as prokaryotes, from the Greek "*pro-karyon*" (before core) due to the lack of a nucleus, successfully inhabit all oceanic habitats. Prokaryotes are found from several hundred metres below the seafloor (Schippers *et al.*, 2005) to the ocean surface (Franklin *et al.*, 2005), in super-cold brine channels of icebergs and sea-ice (Margesina & Miteva, 2011) and hot waters originating in hydrothermal vents (Nakagawa *et al.*, 2004).

Prokaryotes constitute the largest fraction of the living biomass of the world's oceans (Whitman *et al.*, 1998) and play a key role in all biogeochemical cycles in the ocean (Azam & Worden, 2004; Kirchman, 2008). However, our knowledge on their distribution patterns and functional response throughout the ocean, their response to different environmental parameters, and their ecological roles, is still rather limited.

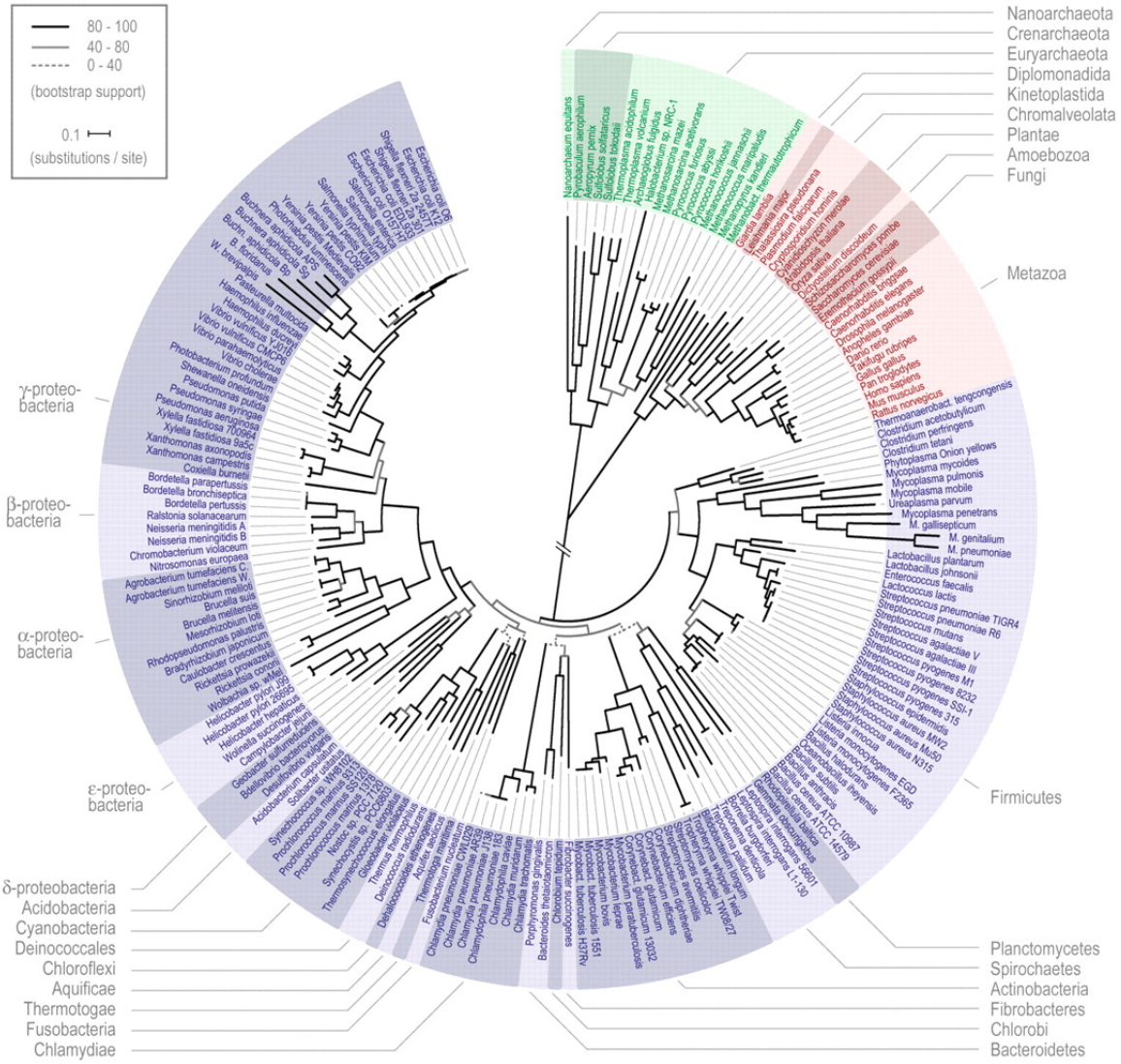

Figure 1.2: The Tree of life (according to Ciccarelli *et al.*, 2006). In clockwise direction and starting from the top, Archaea (Green), Eukaryota (Red) and Bacteria (Blue). Color shadings indicate subdivisions. The branch separating Eukaryota and Archaea from Bacteria has been shortened for display purposes.

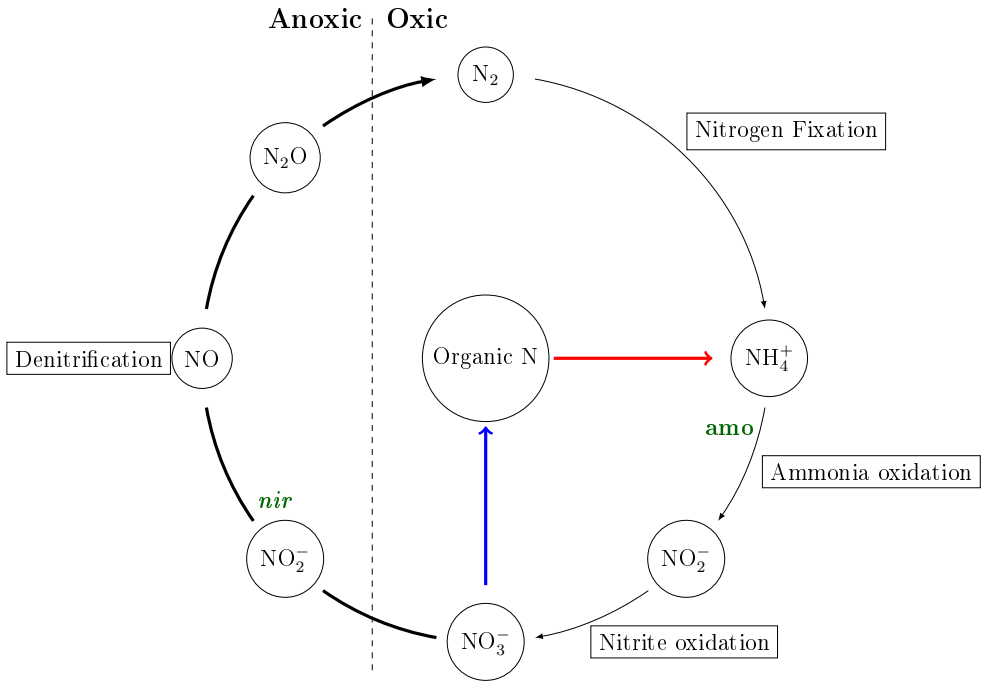

Figure 1.3: Simplified scheme of the nitrogen cycle. Arrows represent different processes (labeled accordingly), with $\longrightarrow$ representing mineralization/ammonification and $\longrightarrow$ representing assimilation. Anoxic and oxic environments are separated by the vertical dashed line. Key prokaryotic genes involved in these processes and discussed in the text are shown in green in the inner side of the circle.

1.3 The Nitrogen Cycle

Nitrogen (N) is present in all living organisms, found in nucleic acids and amino acids (the building blocks for DNA-RNA and proteins, respectively) and other biomolecules. N is transformed through different chemical forms in the nitrogen cycle (a simplified scheme is shown in Figure 1.3). Atmospheric N_2 is reduced to ammonium (NH_4^+)via nitrogen fixation and can be afterwards incorporated into organic molecules, process known as assimilation. Ammonium can also subsequently be oxidized to nitrite (NO_2^-), a process known as the ammonia oxidation. NO_2^- can be further oxidized to nitrate (NO_3^-) via nitrite oxidation. Nitrate, in a series of reduction steps normally under anoxic or hypoxic conditions (see Figure 1.3, left side of dashed line) can be transformed back to N_2, in a process named denitrification. The process by which organic nitrogen, *i.e.*, incorporated in molecules with carbon, is turned into inorganic nitrogen in the form of ammonium, is known as mineralization or ammonification. All the different processes from the N cycle are mediated by Bacteria and to a lesser extent Archaea. Nitrogen fixation, nitrification and denitrification are performed by prokaryotes. However, Archaea have not been shown to perform nitrite oxidation (Madigan *et al.*, 2012).

1.4 Prokaryotic diversity

Microbiologists are now using molecular tools to estimate diversity and identify microbes which do not have cultured representatives. One of the tools used is ortholog genes, *i.e.*, genes that have changed in sequence over time as species have diverged (Madigan *et al.*, 2012). Organisms that share the same orthologs are considered a phylotype. Genes encoding the small subunit of the ribosomal RNA (rRNAs), in particular the *16S rRNA*, are widely used to reconstruct the phylogeny of microorganisms as it is highly conserved (Weisburg *et al.*, 1991).

Using these genetic tools, we can identify microbial operational taxonomic units(OTUs) and we can conduct studies on microbial biogeography, diversity, *etc.* Prokaryotes, similarly to eukaryotes, show macroecological distribution patterns, such as latitudinal species richness gradients (Fuhrman *et al.*, 2008; Pom-

mier *et al.*, 2007), the Rapoport rule, *i.e.*, latitudinal ranges of organisms are generally smaller at lower latitudes than at higher latitudes (Amend *et al.*, 2013; Sul *et al.*, 2013). Contrasting patterns of microbes as compared to macroorganisms, such as no elevation gradient in diversity and distinct generalist/specialist spatial distribution (microbial specialists with broader distribution than generalists) (Carbonero *et al.*, 2014; Fierer *et al.*, 2011). Therefore, the existence of microbial biogeography is now widely accepted (Parnell *et al.*, 2010; Sintes *et al.*, 2015; Sul *et al.*, 2013). However, the underlying factors behind these biogeographical patterns are still poorly characterized. Whether the distribution of microbial OTUs is depending on competition and follows the classical niche theory or, as proposed by the neutral theory, is due to dispersal and drift is still under debate (Vellend *et al.*, 2014).

1.5 Functional diversity

Phylogenetic diversity does not provide information on the ecosystem functioning. All organisms perform metabolic processes mediated by proteins that are encoded by specific genes. These functional genes can be used to assess the distribution and diversity of the organisms according to their potential function in the ecosystem. Ammonia monooxygenase, encoded by the *amo* gene, responsible for ammonia oxidation, and nitrite reductase, encoded by the *nir* gene (Figure 1.3), metabolizing NO_2^- into NO, a process known as nitrite reduction, are two key enzymes in the nitrogen cycle. The abundance and diversity of functional genes can subsequently be assessed using molecular techniques, such as quantitative PCR and sequencing.

1.6 Modelling

According to the Oxford Dictionary (Stevenson, 2015): "Model - A. Noun. 8. A simplified or idealized description or conception of a particular system, situation, or process, often in mathematical terms, that is put forward as a basis for theoretical or empirical understanding, or for calculations, predictions, *etc.*; a conceptual or mental representation of something.".

Models have been developed to explain the mechanisms behind diverse phenomena. They have been used to explain the movement of celestial bodies with Newton laws (Newton, 1687), quantum mechanics (Schrödinger, 1926), bacterial growth (Zwietering *et al.*, 1990), evolution (Hanage *et al.*, 2006) and even life (artificial life) models (Gardner, 1970). Thus, a model should provide simulations without having to rely on future observations of the target of the study. The usual representation of a model is based on simple equations, such as regression models or the Newton's law of universal gravitation (Newton, 1687):

$$F = G\frac{m_1 m_2}{r^2} \tag{1.1}$$

Alternatively, algorithms can be used.

1.7 Objectives

In the present study, we aim to (1) characterize the geographical distribution of both Archaea and Bacteria from pole-to-pole across the Atlantic Ocean, (2) to assess the relative abundance and diversity of different prokaryotic genes related to specific metabolic pathways from the nitrogen cycle, and finally, by using the information obtained from the previous two objectives, (3) to build a model based on the environmental conditions to predict the distribution and diversity of Archaea and Bacteria in the ocean, both at a functional and at a phylogenetic level. In order to achieve these objectives, high throughput sequencing of the *16S rRNA* and the *nir*K gene from Archaea and Bacteria was carried out using DNA extraction from water samples obtained from different depth layers along a latitudinal transect in the Atlantic Ocean.

Chapter 2

Materials and Methods

2.1 Sampling

Seawater samples were collected during the GEOTRACES-1 and -2 cruises onboard R/V *Pelagia*, from April to June 2010, and GEOTRACES-3 on board of R/V *James Cook* from February to April 2011. Samples were collected at 6-8 depths in 51 stations from 65°N to 55°S (Figure 2.1) from the epipelagic [0 - 199 m], mesopelagic [200 - 999 m], upper bathypelagic [1000 - 1999 m] and lower bathypelagic [>2000 m] depths. Water samples were collected with Niskin bottles mounted in a frame holding also sensors for conductivity, temperature, depth (CTD), salinity, oxygen, fluorescence and optical backscattering. Six different oceanographic regions were distinguished according to Longhurst (Longhurst, 2007, from north to south: North Atlantic Arctic province (ARCT), North Atlantic Drift province (NADR), North Atlantic Gyral province (NAG), Western Tropical Atlantic (WTRA), South Atlantic Gyral (SATL) and Subantarctic province (SANT); Table 2.1 and Figure 2.1). The methods used for the measurement of inorganic nutrients and trace elements are available at the Geotraces website (`http://www.geotraces.org`) (Group *et al.*, 2015). Bacterial abundance was assessed by flow cytometry as previously described (Sintes *et al.*, 2015). Inorganic nutrients (SiO_4^{4-}, PO_4^{3-}, NO_3^-, NO_2^-) and trace elements (Al, Cd, Fe, Mn, Ni, Pb, Zn, Y, La), and bacterial activity (^{3}H-Leucine uptake) were measured (Table 2.2) at 24 depth layers. Samples for prokaryotic activity measurements (leucine uptake and dissolved inorganic carbon (DIC) fixation)

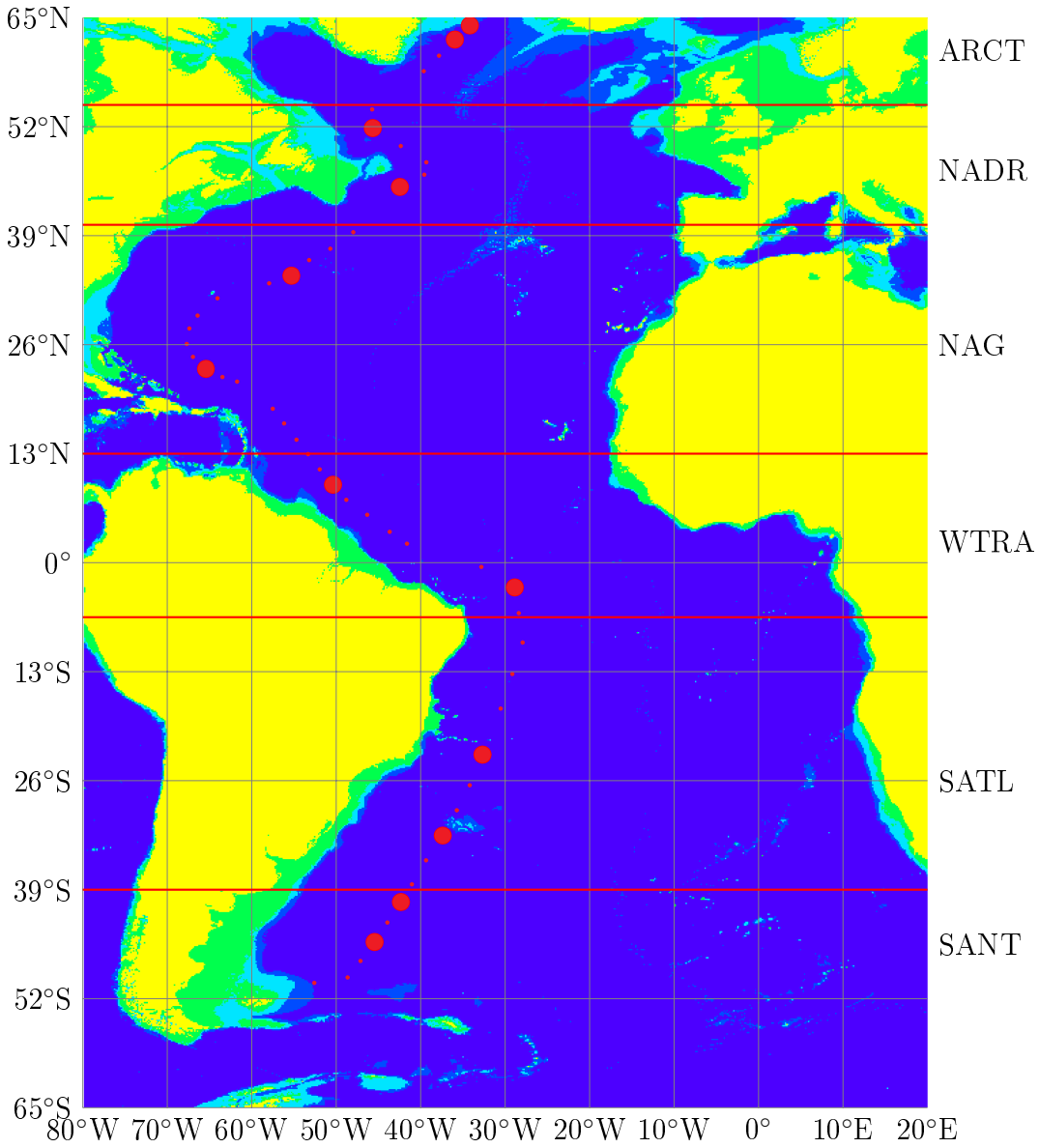

Figure 2.1: Location of the sampling sites along the cruise track. Landmasses are colored in , 0 - 199 m depth in , 200 - 999 m depth in , 1000 - 1999 m depth in and >2000 m depth in . Sampling sites appear as •, while stations where biological samples for sequencing of 16S rRNA and *nir*K genes were collected are represented by ●. Oceanographic regions names appear on the right, with division lines in red. Bathymetric data retrieved from the General Bathymetric Chart of the Oceans website(`http://www.gebco.net/`, 2014)

Table 2.1: Number of samples in each ocean province (Longhurst, 2007) and water layer. The North Atlantic Arctic province (ARCT; 70°N–55°N), the North Atlantic Drift province (NADR; 55°N–40°N), the North Atlantic Gyral province (NAG [comprising the North Atlantic Tropical and the Subtropical Gyral province (40°N–12°N)]), the Western Tropical Atlantic (WTRA; 12°N–6°S) province, the South Atlantic Gyral (SATL; 6°S–40°S) and the Subantarctic province (SANT [comprising the Subtropical Convergence Zone (40°S–45°S) and the Subantarctic Water Ring province (45°S–55°S)]). Epipelagic(0 - 199 m), Mesopelagic(200 - 999 m), Upper Bathypelagic(1000 - 1999 m), Lower Bathypelagic(>2000 m).

			Bathipelagic	
	Epipelagic	Mesopelagic	Upper	Lower
ARCT	2	2	2	2
NADR	2	2	2	2
NAG	2	3	3	3
WTRA	2	2	2	2
SATL	2	2	2	2
SANT	2	2	2	2

and DNA extraction were collected at 6-8 depths. Prokaryotic leucine uptake (De Corte *et al.*, 2016) and DIC fixation (Herndl *et al.*, 2005) were evaluated via the incorporation of radiolabeled substrates as previously described. Samples from four depths obtained from two stations per region were used to characterize the *nir*K and *16S rRNA* gene of Bacteria and Archaea (Table 2.1, Figure 2.1) by Illumina sequencing.

2.2 DNA extraction

Two-10 L of seawater were filtered onto 0.22μm polycarbonate filters depending on the depth. The filters were stored at -80°C until DNA extraction was performed using Ultraclean soil DNA isolation kit (MoBIO) at the facilities of the Department of Limnology and Bio-Oceanography, University of Vienna.

Table 2.2: Nature of the Data collected at the Geotraces cruises and used in this study.

Environmental Parameter	Unit
Si	μmol/kg
PO_4^{3-}	μmol/kg
NO_3^-	μmol/kg
NO_2^-	μmol/kg
Al	nmol/kg
Cd	nmol/kg
Fe	nmol/kg
Mn	nmol/kg
Ni	nmol/kg
Pb	pmol/kg
Zn	nmol/kg
Y	pmol/kg
La	pmol/kg
^{3}H-Leucine uptake	pmol/L/d
Temperature	°C
Depth	m
Salinity	-
Latitude	Absolute °
Oxygen	μmol/kg
Fluorescence	arb
TALK	μmol/kg

2.3 PCR and Sequencing

The 16S rRNA gene from Bacteria and Archaea, and the *nir*K gene from Bacteria and Archaea were PCR amplified using the primers described in Table 2.3 prior to sequencing at the facilities of the Department of Limnology and Bio-Oceanography, University of Vienna. Each 25 μL PCR reaction consisted of 0.2-1.0 μM of the corresponding primers (Table 2.3), 200 μM of dNTP, 2μg BSA, 2.5 mM $MgCl_2$, and 2.5U Picomaxx high fidelity DNA polymerase (Agilent Technologies), 2.5 μL of the corresponding PCR buffer, and 1-2 μL of the DNA extract, made up to 25 μL with UV-treated ultra-pure water (Sigma). Cycling conditions for bacterial *16S rRNA* gene were as follows: 5 min at 94°C, followed by 25 cycles consisting at 94°C for 30 sec, 57.5°C for 30 sec and 72°C for 45 sec.

Cycling was followed by a final amplification step at 72°C for 10 min and then held at 4°C. The thermocyling for archaeal *16S rRNA* gene consisted of an initial denaturation step at 95°C for 5 min, followed by 10 touch-up PCR cycles consisting of 30 sec at 94°C, increasing annealing temperature from 50°C to 54°C (increasing 0.5°C each cycle) for 40 sec, 72°C for 1 min, 20 cycles of 94°C for 30 sec, 54°C for 40 sec, 72°C for 1 min, followed by one cycle of 72°C for 10 min and 4°C hold.

Bacterial *nir*K gene thermal cycling consisted of an initial denaturation step at 95°C for 5 min, 9 touchdown cycles at 95°C for 30 sec, decreasing annealing temperature from 68°C to 60°C for 30 sec (progressively decreased by 1°C), and 81.5°C for 60 sec. This was followed by 26 cycles at 95°C for 30 sec, 60°C for 30 sec, at 81.5°C for 60 sec, and a final extension step at 72°C for 10 min. Archaeal *nir*K -a and -b were amplified by an initial denaturation at 94°C for 4 min, followed by 30 cycles consisting of denaturation at 94°C for 60 sec, annealing at 50°C for 60 sec, an extension at 72°C for 60 sec, and a final extension step at 72°C for 10 min.

After the PCR, the different products were checked on a 2% agarose gel for the correct band size, and the PCR product was purified using PCRExtract MiniKit (5-PRIME). Purified PCR products were quantified using a Nanodrop® spectrophotometer. Subsequently, all PCR products were standardized to a concentration of 20 ng DNA μL^{-1} with PCR-grade water (Sigma) prior to sequencing.

Sequencing was performed on the Illumina MiSeq ® next generation sequencing system (Illumina Inc.). The resulting 2 x 300 bp reads were demultiplexed.

The final output from three different pipelines (Uparse, Qiime, Mothur) were compared. Specifically, the total number of estimated OTUs present and the diversity indexes (Shannon, Shannon evenness and Chao) obtained with the three different pipelines were compared.

The results obtained with the three pipelines were similar (see results section). Thus, we subsequently used the Uparse based pipeline. This pipeline mainly used usearch tools and commands, as the Uparse clustering tool is inserted in this software collection. Therefore, almost all (except for the taxonomy classification) the steps were conducted with usearch. Both forward and reverse reads were merged, followed by filtering (minimum length of 400 nucleotides, no n's, no reads with more than 4 total expected errors). The merged and filtered reads were then dereplicated (a list of unique sequences was kept in this step, with the number of repeated sequences annotated to the sequences labels) sorted (with all singletons discarded), and clustered into different OTU's, using the Uparse commands. fasta_number (python script) was used to label the sequences of each OTU, with its respective OTU code, and with this output, all the reads were mapped against. The previous mapping file was used as an input for uc2otutab (python script) to create the OTU table. To access the taxonomy to the OTUs, a blast was carried with the labelled OTU sequences against the SILVA_123 database (for *16S rRNA* gene) and a in house *nir*K database (for the *nir*K gene).

Table 2.3: Primer sets and PCR conditions used for the different genes.

Target	Analysis	Primer name	Primer sequence
Bacteria	qPCR	recAF	TGTGCITTTATWGATGCIGAGCATGC
*rec*A		recAR	CCCATGTCICCTTCKATTTCIGCTTT
Bacteria	qPCR/	nirKq-F	TCATGGTGCTGCCGCGYGA
*nir*K	Sequencing	nirK1040	GCCTCGATCAGRTTRTGGTT
Bacteria	Sequencing	Bakt-341F	CCTACGGGNGGCWGCAG
16S		Bakt-805R	GACTACHVGGGTATCTAATCC
Archaea	qPCR	GI-751F	GTCTACCAGAACAYGTTC
16S		GI-956R	HGGCGTTGACTCCAATTG
Archaea	qPCR/	anirKa-61F	ACBYTATTCGGAAGYACATACACA
*nir*K-a	Sequencing	anirKa-579R	GYMATTCCGTACATKCCGGA
Archaea	qPCR/	anirKb-58F	CTATTCGGARGTWCTTTYACTGC
*nir*K-b	Sequencing	anirKb-555R	ACGTGTTGGTCCATTGCTGC
Archaea	Sequencing	Arch349F	GYGCASCAGKCGMGAAW
16S		Arch806R	GGACTACVSGGGTATCTAAT

continues

continuation

Target	Annealing	Primer concentration (μM)	Fragment (bp)	Reference
Bacteria *rec*A	53	0.5	212	(Holmes *et al.*, 2004)
Bacteria *nir*K	68-60	0.75	472	(Mosier & Francis, 2010)
Bacteria 16S	57.5	0.5	465	(Klindworth *et al.*, 2012)
Archaea 16S	58	0.2	205	(Mincer *et al.*, 2007)
Archaea *nir*K-a	50	1.0	518	(Lund *et al.*, 2012)
Archaea *nir*K-b	50	1.0	497	(Lund *et al.*, 2012)
Archaea 16S	50-54	0.5	458	(Takai & Horikoshi, 2000)

2.4 Modelling and Statistics

The developed model (Annex 1) calculated for each OTU the multidimensional normal distribution in relation to selected environmental data (absolute latitude, depth, temperature, concentrations of Ni, Fe, Zn, Mn, Si, PO_4, NO_2, bacterial abundance, leucine incorporation). For this, the averages of the samples were calculated for each environmental variable, where each observation was weighted according to OTU frequency detected on that same sample,

$$\bar{E}_{x,otu} = \frac{\sum_{i=1}^{i=n} e_{i,x,otu} w_{i,x,otu}}{\sum_{i=1}^{i=n} w_{i,x,otu}} \tag{2.1}$$

where $\bar{E}$ is the Environmental variable x on OTU otu, e is the observation, i is the sample and w is the weight(frequency). For the calculation of the distance between this multidimensional point of averages, and a sample, it was used the Mahalanobis distance (Mahalanobis, 1936),

$$M = \sqrt{(u - v_i)V^{-1}(u - v_i)^T} \tag{2.2}$$

Where M is the Mahalanobis distance, V^{-1} is the inverse of the covariance matrix, u and v_i are two arrays of values(u with the averages of $\bar{E}$, and v_i with the sample i environmental values) and T stands for transposed. The Mahalanobis distance was used as a proxy for standard deviation. Further down, the abundance of a specific OTU was calculated for each sample as so,

$$Mod_{i,otu} = \bar{f}_{otu} \pm M_i \times sd_{otu} \tag{2.3}$$

Where Mod is the modeled value (of abundance, *i.e.* frequency) for the sample, M is the Mahalanobis distance, i is the sample and both $\bar{f}_{otu}$ and sd_{otu} are the average and the standard deviation of the OTU otu frequency, respectively. From this equation, two results ($-$ and $+$) are compared to the original value of (frequency) abundance, with the closest to the original being kept. All values below 0 were turned into 0. This was applied to all OTUs throughout the entire sampling locations.

The diversity indexes were calculated from the estimated OTUs and compared to the diversity indexes estimated from the sequences obtained in the different samples. All the simulations and algorithms were run and written using python programming language (Rossum & Drake Jr., 2006).

All statistical analyses used a significance level of $\alpha = 0.05$, and were performed using R (R Development Core Team, 2008), packages (Chen, 2012; Dray *et al.*, 2007; Hijmans & Van Etten, 2014; Lemon, 2006; Paradis *et al.*, 2004; Pierce, 2012; Schliep, 2011).

Chapter 3

Results

3.1 Comparison of different pipelines for analysis of high throughput sequencing data

16S rRNA gene and *nir*K gene sequencing of Bacteria and Archaea from the Atlantic resulted in a total of 4111873 sequences (Table 3.1), with a majority of archaeal sequences (>3.5 million 16S rRNA gene sequences from Archaea).

We analysed the obtained sequences with three different available pipelines: Uparse, Qiime, and Mothur. These 3 pipelines resulted in 1986, 7941 and 244734 OTUs for the whole dataset of Bacteria (Table 3.2).

The processing time with the three pipelines was exponentially increasing from Uparse to Qiime and to Mothur, lasting approximately 3 h, 2 d and 2-3 weeks, respectively. The patterns of diversity and community composition obtained from the different pipelines were similar (Table 3.2). However, the diversity and richness indexes were UPARSE <QIIME <MOTHUR, in agreement with previous reports indicating that both Mothur and Qiime might overestimate

Table 3.1: Samples recovered from the Geotraces campaign 1,2 and 3 (after merging forward and reverse pair and filtering[read minimum length:400; maximum N's:0; Maximum expected number of errors: 4]).

16S Archaea	16S Bacteria	*nir*K Bacteria	*nir*K-a Archaea	*nir*K-b Archaea
3557904	204816	230630	931512	464441

Table 3.2: Comparison of the results of each pipeline on the 16S Bacteria dataset. Average diversity indexes and total number of OTUs.

Pipeline	Shannon	Shannon Eveness	Chao	Total Number of OTUs
Uparse	3.64	0.76	208	1881
Qiime	4.82	0.85	1039	7941
Mothur	6.61	0.79	9467	244734

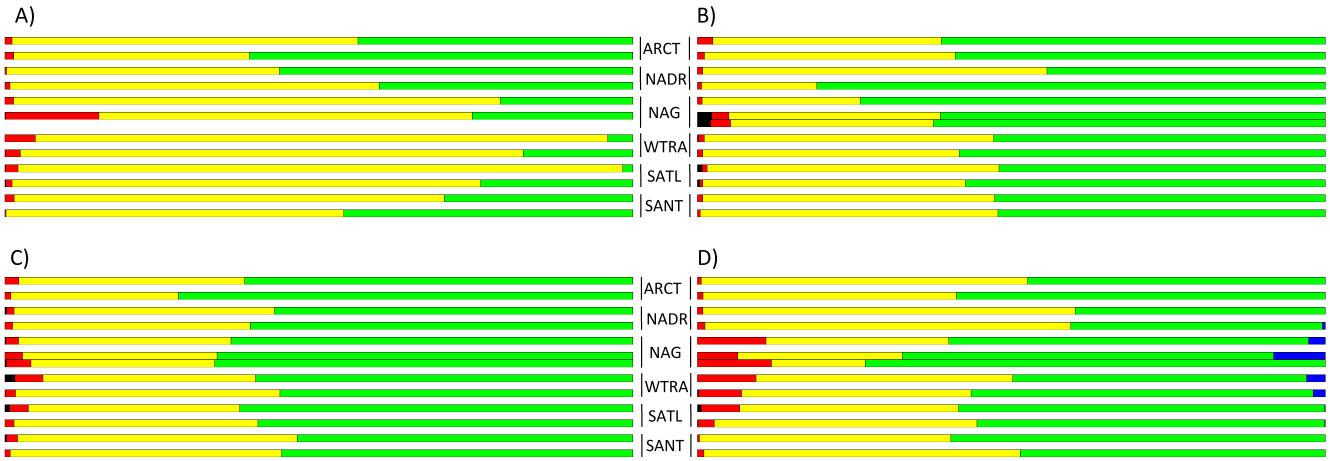

Figure 3.1: Distribution of main phyla taxa of Archaea through the Atlantic ocean. X-axis: relative abundance (%) of the taxa (a) in the epipelagic (b) mesopelagic (c) upper bathypelagic and (d) lower bathypelagic realms. From left to right in the barplots: ■ - Aigarchaeota, ■ - Woesearchaeota, ■ - Euryarchaeota, ■ - Thaumarchaeota, ■ - Marine Hidrothermal Vent Group; Y-axis: name of the sample and relative latitude.

the number of OTUs (Majaneva *et al.*, 2015). Thus, we subsequently focused on the results obtained using the Uparse pipeline.

3.2 Phylogenetic characterization of Bacteria and Archaea based on *16S rRNA gene*

3.2.1 16S

Archaeal phyla detected in the Atlantic included Euryarchaeota, Woesearchaeota (DHVEG-6) and Thaumarchaeota as the most abundant groups. Other lower

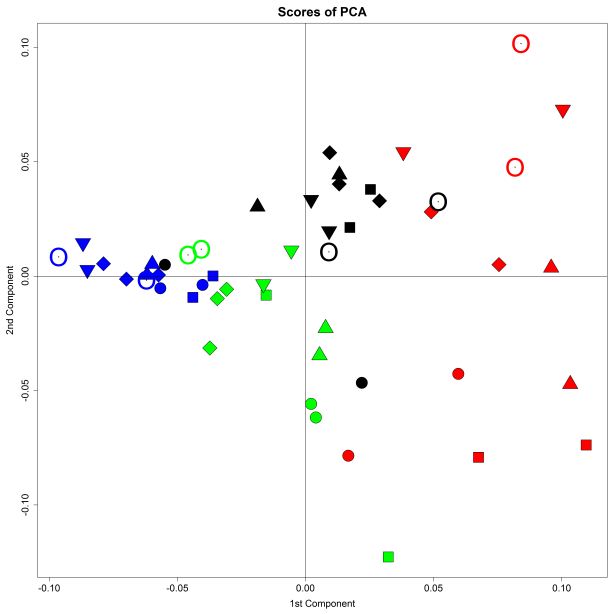

Figure 3.2: Principal Coordinates analysis of the 16S rRNA gene of Archaea. - Epipelagic, - Mesopelagic, - Upper Bathypelagic, - Lower Bathypelagic. ● - ARCT, ■ - NADR, ◆ - NAG, O - WTRA, ▼ - SATL, ▲ - SANT.

abundance phyla comprised Aigarchaeota and the marine hydrothermal vent group (MHVG).

The dominant phylum was Thaumarchaeota with an average of 51.7 % of the archaeal OTUs, followed by Euryarchaeota with an average 45.2 % of the OTUs.

Euryarchaeota increased in their relative contribution to the archaeal communities towards lower latitudes in the epipelagic environment (Figure 3.1.a) in contrast to Thaumarchaeota. In contrast, Euryarchaeota inhabiting the lower bathypelagic realm decreased in their relative abundance from high to low latitudes. Thaumarchaeota dominated the archaeal community in mesopelagic waters (Figures 3.1.b, c, d), contributing up to 74.0 % and 81.0 % of archaeal sequences in the NAG and NADR, respectively.

Woesearchaeota (DHVEG-6) contributed on average 2.6 % to the archaeal communities, increasing in their relative abundance from high towards lower latitudes in all depth layers, except in the mesopelagic environment. Woesearchaeota were mostly uniformly distributed in mesopelagic waters, with slightly higher abundance at the NAG.

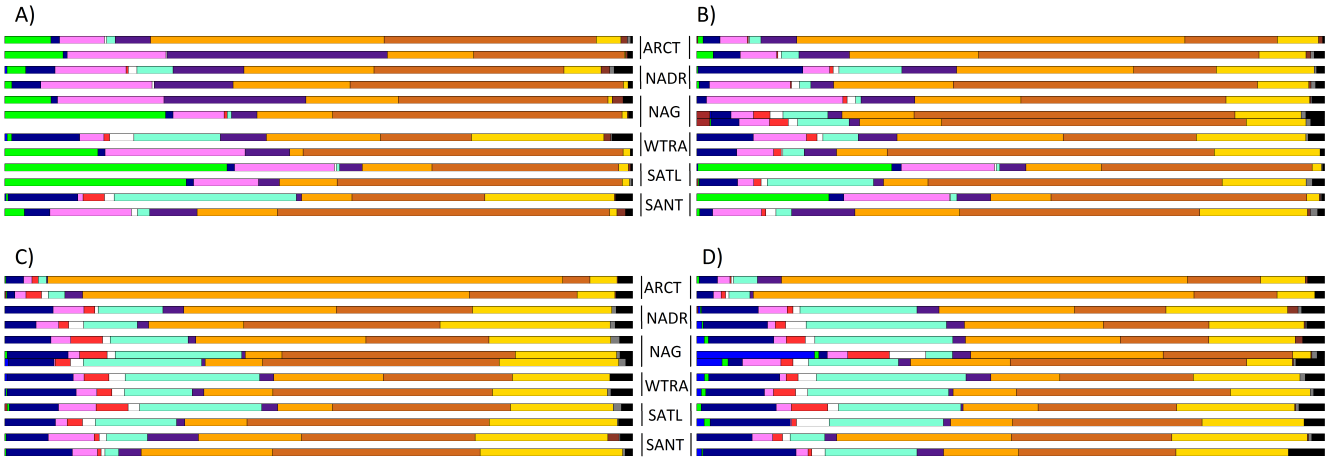

Figure 3.3: Distribution of main phyla of Bacteria throughout the Atlantic ocean. X-axis: relative abundance (%) of the taxa (a) in the epipelagic (b) mesopelagic (c) upper bathypelagic and (d) lower bathypelagic realms. The phylum Proteobacteria was divided into its main classes (α, γ, δ, β and others), while the low abundance phyla were combined together into the group 'others'. From left to right in the barplots: - Nitrospirae, - Parcubacteria, - Cyanobacteria, - Marinimicrobia (SAR406), - Actinobacteria, - Acidobacteria, - Planctomycetes, - Chloroflexi, - Bacteroidetes, - Gammaproteobacteria, - Alphaproteobacteria, - Deltaproteobacteria, - Betaproteobacteria, - other Proteobacteria, - Others; Y-axis: name of the sample and relative latitude.

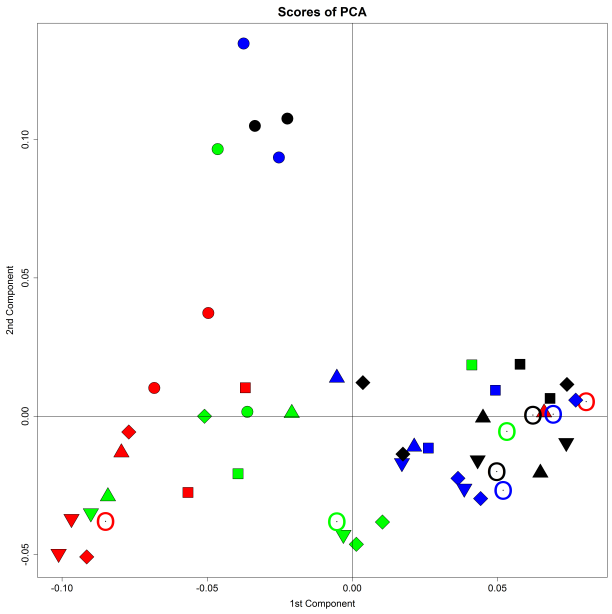

Figure 3.4: Principal Coordinates analysis of the 16S rRNA gene of Bacteria. ■ - Epipelagic, ■ - Mesopelagic, ■ - Upper Bathypelagic, ■ - Lower Bathypelagic. ● - ARCT, ■ - NADR, ◆ - NAG, O - WTRA, ▼ - SATL, ▲ - SANT.

Members of the MHVG were limited to the lower bathypelagic layers, where they represented ∼ 1.3 % of the archaeal communities in the NADR, NAG, WTRA and SATL (Figure 3.1.d).

Aigarchaeota, with an average contribution of 0.2 % to the archaeal communities, were mainly found in the communities below the epipelagic zone, with a maximum at the southern station of the NAG in mesopelagic waters.

Archaeal communities throughout the Atlantic were stratified (Figure 3.2). Archaeal communities from epipelagic waters clustered together and were closer related to mesopelagic communities than to those from deeper layers. Likewise, mesopelagic communities clustered closer to upper bathypelagic than to lower bathypelagic communities. Interestingly, epipelagic and lower bathypelagic communities were closely represented in the PCoA plot, indicating a higher similarity between these two communities as compared to the upper bathypelagic (Figure 3.2). Epi- and mesopelagic archaeal communities showed a larger dissimilarity between samples from different locations (Figure 3.2) as compared to the bathypelagic communities.

The dominant bacterial phyla included Proteobacteria, Cyanobacteria, Actinobacteria, Chloroflexi, Bacteroidetes, Acidobacteria, Planctomycetes and Marinimicrobia (SAR406). Other phyla present were Nitrospirae, Parcubacteria, Verrucomicrobia, LCP-89, Lentisphaerae, Hydrogenedentes, Deinococcus-Thermus, Gracilibacteria, Spirochaetes, PAUC34f, Gemmatimonadetes, Sacharibacteria, Deferribacteres and Firmicutes.

The most abundant phylum was Proteobacteria, accounting on average 63.7 % of the bacterial community. Alpha-, Gamma-, Delta- and Betaproteobacteria represented on average 30.1 %, 21.3 %, 11.6 % and 0.3 % of the bacterial community, respectively.

The Chloroflexi phylum constituted on average 8.9 % of the bacterial OTUs. Chloroflexi increased towards the bathypelagic layers, where they accounted for 11.4 % and 14.8 % of the upper and lower bathypelagic bacterial OTUs, respectively. Moreover, the relative contribution of Chloroflexi increased towards low latitudes at all depth layers (Figures 3.3.c, d). Actinobacteria, with an average contribution of 6.5 % to the bacterial communities, was relatively more abundant in epipelagic and mesopelagic layers. In epipelagic waters, Actinobacteria were less abundant in SANT, WTRA and ARCT regions, with relative abundances ranging between 0.9 and 7.3 % of the bacterial community. In mesopelagic waters from ARCT, NAG and SATL, Actinobacteria ranged between 3.0 % and 5.8% of the bacterial community (Figure 3.3.b). SAR406 account for 6 % of the bacterial communities. SAR406 members increased in their relative abundance with depth, with average contribution ranging between 14.4 % and 35.4 %, from epi- to lower bathypelagic waters. The relative abundance of Bacteroidetes decreased with depth from 10.2 % in epipelagic waters to 2.3 % in lower bathypelagic environments. In epipelagic waters Bacteroidetes the highest contribution to bacterial communities was detected in the northern hemisphere, while Bacteroidetes inhabiting deeper layers were more abundant at high latitudes, both in the north and south (Figure 3.3). Cyanobacteria were present in the epipelagic layer at a high relative abundance, on average 11.6 % of the bacterial OTUs, and to a lesser extent in the mesopelagic realm mainly in SATL and SANT, with an average of 4.8 %.

Acidobacteria represented on average 1.6 % of the bacterial communities. The highest abundance of Acidobacteria was found in the bathypelagic realm, with an average contribution of 2.6 % to bathypelagic bacterial communities (Figure 3.3).

Planctomycetes, with an average contribution of 1.4 % to the bacterial community, increased its contribution with depth and reached up to 2.3 % in lower bathypelagic communities.

The relative abundance of Nitrospirae and Parcubacteria peaked at NAG, where Nitrospirae contributed up to 2.1 % in the mesopelagic and Parcubacteria contributed 18.7 % in the lower bathypelagic waters (Figures 3.3.b, d).

Similarly to archaeal communities, bacterial communities were depth stratified, with higher similarities between communities from specific depth layers as compared to communities from specific regions. However, ARCT and WTRA communities formed a separate cluster of deep ocean bacterial communities (Figure 3.4).

3.2.2 Nitrite reductase containing Bacteria and Archaea

Archaeal cells harbouring the *nir*K gene were divided in two groups (a and b) according to their sequences, Archaeal *nir*K-a containing communities consisted of three main clusters. Upper bathypelagic communities from throughout the Atlantic clustered together with the ARCT communities from all depth layers and SANT communities from the mesopelagic. Lower bathypelagic communities formed a separate cluster and meso- and epipelagic communities from other regions except the ARCT (and SANT) region formed the third cluster (Figure 3.5).

Archaeal *nir*K-b communities were composed of one well defined group comprising the upper bathypelagic communities and the ARCT lower bathypelagic, and a second group which included the lower bathypelagic communities. The rest of the archaeal *nir*K-b harbouring communities from meso and epipelagic realms were scattered (Figure 3.6).

Bacteria harbouring *nir*K clustered mainly according to the oceanographic region, with ARCT and NAG communities forming separate clusters, and lower

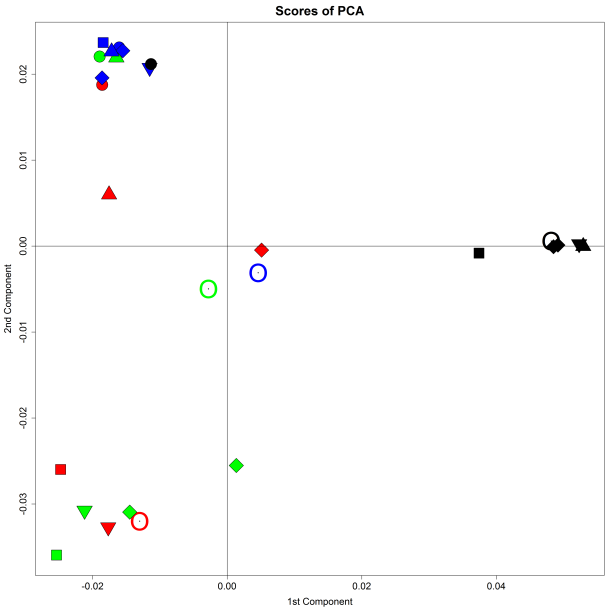

Figure 3.5: Principal Coordinates analysis of the *nir*K-a gene of Archaea. ■ - Epipelagic, ■ - Mesopelagic, ■ - Upper Bathypelagic, ■ - Lower Bathypelagic. ● - ARCT, ■ - NADR, ◆ - NAG, O - WTRA, ▼ - SATL, ▲ - SANT.

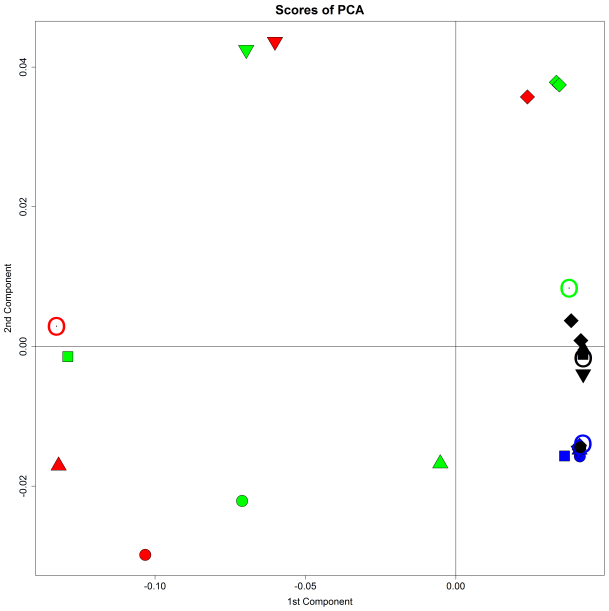

Figure 3.6: Principal Coordinates analysis of the *nir*K-b gene of Archaea. ■ - Epipelagic, ■ - Mesopelagic, ■ - Upper Bathypelagic, ■ - Lower Bathypelagic. ● - ARCT, ■ - NADR, ◆ - NAG, O - WTRA, ▼ - SATL, ▲ - SANT.

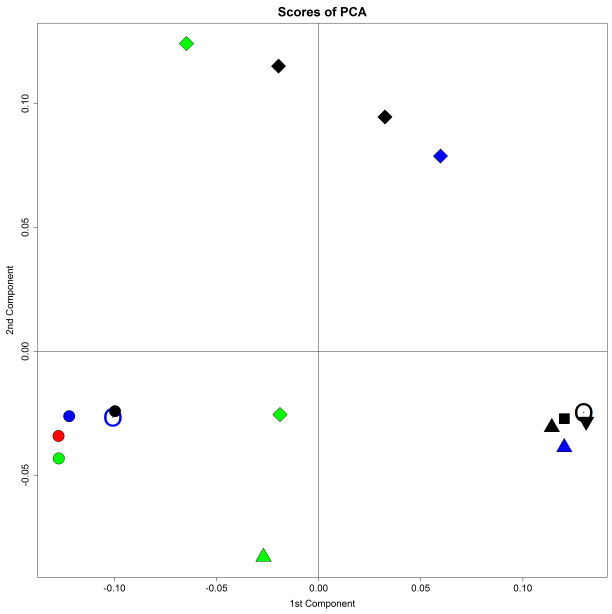

Figure 3.7: Principal Coordinates analysis of the *nir*K gene of Bacteria. ■ - Epipelagic, ■ - Mesopelagic, ■ - Upper Bathypelagic, ■ - Lower Bathypelagic. ● - ARCT, ■ - NADR, ◆ - NAG, O - WTRA, ▼ - SATL, ▲ - SANT.

bathypelagic communities constituting a third separate cluster (Figure 3.7).

3.3 Distribution of bacterial and archaeal *nir*K harbouring communities

3.3.1 Diversity of bacterial and archaeal communities

The diversity, evenness and richness indexes were in general higher for Bacteria (Table 3.4) than for Archaea (Table 3.3) (Shannon diversity, Mann-Whitney: W= 131, p-value<0.001; Shannon Evenness, Mann-Whitney: W= 270, p-value<0.001; Chao, Mann-Whitney: W= 18, p-value<0.001).

In relation to depth, the Chao richness index of Archaea (Table 3.3) was lower (Kruskall-Wallis: χ^2: 12.84, p-value=0.005) in epipelagic and lower bathypelagic communities as compared to upper bathypelagic (Multiple comparison test after Kruskal-Wallis; p-value<0.05). In contrast, the Chao index of Bacteria (Table 3.4) was higher (Kruskall-Wallis: χ^2: 9.26, p-value=0.026) for deeper

Table 3.3: Diversity indexes (mean±sd) for Archaea 16S throughout the sampling regions and water layers. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	2.45±0.16	2.45±0.28	2.58±0.07	2.28±0.08	2.54±0.36	2.12±0.68
Meso	2.34±0.19	2.59±0.3	2.63±0.49	2.38±0.02	2.63±0.1	2.63±0.14
UB	2.55±0.15	2.8±0.07	2.65±0.09	2.53±0.38	2.35±0.09	2.44±0.08
LB	2.59±0.27	2.17±0.23	2.51±0.1	2.97±0.04	2.78±0.2	1.89±0.01

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.62±0	0.61±0.01	0.66±0.06	0.62±0.02	0.65±0.02	0.59±0.17
Meso	0.62±0.06	0.62±0.04	0.63±0.11	0.57±0.01	0.64±0.02	0.65±0.01
UB	0.63±0.06	0.67±0.02	0.63±0.01	0.6±0.08	0.58±0.01	0.58±0.02
LB	0.64±0.07	0.55±0.03	0.75±0.06	0.7±0.01	0.66±0.06	0.54±0.07

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	61±17	67±23	61±23	59±16	65±15	54±10
Meso	69±0	92±46	89±28	82±6	71±9	69±17
UB	69±20	79±3	79±11	87±10	81±5	75±10
LB	67±13	65±13	35±15	75±9	88±21	43±13

layers (mesopelagic and bathypelagic) as compared to the epipelagic realm (Multiple comparison test after Kruskal-Wallis; p-value$<$0.05 and p-value$<$0.10 for epipelagic vs. upper bathypelagic and lower bathypelagic respectively, all other comparisons with p-value$>$0.05). No statistically significant difference was found in diversity, evenness and richness indexes throughout the latitudes in both Bacteria and Archaea (Shannon diversity, p=0.08).

3.3.2 Diversity of *nir*K harbouring communities

The Shannon diversity index of *nir*K-a containing Archaea was lower in the epipelagic and increased towards deeper layers in the NAG,WTRA and SATL. However, the opposite trend was observed in other regions (Table 3.5.a) with the

Table 3.4: Diversity indexes (mean±sd) for Bacteria 16S throughout the sampling sites. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	3.6±0.25	3.76±0.9	3.17±0.38	3.53±0.85	2.57±0.21	3.73±0.62
Meso	3.55±0.28	3.89±0.5	3.88±0.06	3.69±0.44	3.4±0.59	3.4±0.16
UB	2.64±0.46	3.76±0.5	3.94±0.18	4.1±0.57	4.03±0.58	3.67±0.51
LB	2.77±0.64	4.02±0.09	3.87±0.44	4.16±0.13	4.08±0.34	3.71±0.06

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.79±0.02	0.78±0.1	0.71±0.08	0.77±0.11	0.6±0.03	0.78±0.04
Meso	0.76±0.04	0.8±0.06	0.79±0.01	0.79±0.07	0.72±0.08	0.74±0.02
UB	0.59±0.06	0.78±0.06	0.8±0.03	0.82±0.06	0.8±0.08	0.8±0.02
LB	0.61±0.11	0.82±0	0.8±0.07	0.83±0.03	0.81±0.03	0.78±0.03

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	131±36	207±122	143±18	146±73	128±33	214±153
Meso	151±28	233±79	263±51	207±8	206±106	162±29
UB	187±63	227±30	247±71	225±37	301±10	184±141
LB	187±16	277±5	178±39	236±49	300±3	219±10

Table 3.5: Diversity indexes for Archaeal *nir*K-a throughout the sampling sites. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	1.85	1.9	1.09	0.8	0.56	1.58
Meso	0.97	1.9	1.38±0.4	1.82	1.23	1.4
UB	0.48	0.48	1.22±0.2	1.78	1.23	1.39
LB	1.1	1.63	1.1±0.64	1.64	1.83	1.44

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.74	0.86	0.99	0.5	0.4	0.69
Meso	0.54	0.86	0.9±0.14	0.71	0.59	0.64
UB	0.35	0.35	0.72±0.09	0.85	0.59	0.63
LB	0.57	0.74	0.82±0.16	0.84	0.83	0.8

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	26	11	3	8	5	18
Meso	12	6	6±4	25	10	12
UB	5	5	6±3	8	10	24
LB	8	24	8±8	7	19	7

exception of the lower bathypelagic, where diversity was consistently high. Evenness was high in the NAG region (Table 3.5.b). The Chao richness indices were higher in the epipelagic waters of ARCT and SANT, while in other regions, the richness was higher in deeper waters.

*Nir*K-b containing archaeal cells exhibited a higher diversity in mesopelagic and upper bathypelagic environments, especially in the WTRA (Table 3.6.a.c). The evenness of archaeal *nir*K-b followed the same pattern as the diversity (Table 3.6.b). The Chao richness indices were higher in the mesopelagic and upper bathypelagic.

Diversity, evenness and richness indices of *nir*K of Bacteria were higher in the lower bathypelagic waters of the NADR, NAG and SANT, and in the mesopelagic of the SANT. However, *nir*K of Bacteria could not be amplified efficiently in some samples, especially from epipelagic waters (Table 3.7).

Table 3.6: Diversity indexes for Archaeal *nir*K-b throughout the sampling sites. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	2.37	-	3.02	0.25	2.38	0.27
Meso	2.74	1.89	2.23±0.05	3.09	2.37	2.4
UB	2.33	2.46	2.51±0.07	2.34	2.51	2.27
LB	2	0.94	0.78±0.06	1.01	1.3	0.87

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.7	-	0.79	0.18	0.65	0.14
Meso	0.75	0.58	0.63±0.01	0.8	0.63	0.68
UB	0.69	0.67	0.7±0	0.68	0.71	0.64
LB	0.6	0.38	0.34±0.03	0.41	0.46	0.39

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	37	-	49	4	46	10
Meso	42	44	41±5	69	70	40
UB	46	64	49±10	60	40	45
LB	51	13	12±1	12	22	12

Table 3.7: Diversity indexes for Bacterial *nir*K throughout the sampling sites. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.66	-	-	-	-	-
Meso	0.2	-	0.36±0.11	-	-	0.97
UB	0.69	-	0.69	0.13	-	0.2
LB	0.56	2.01	0.78±0.37	0.94	0.66	1.72

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.95	-	-	-	-	-
Meso	0.28	-	0.33±0.1	-	-	0.6
UB	1	-	0.99	0.18	-	0.18
LB	0.81	0.69	0.61±0.2	0.53	0.41	0.67

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	2	-	-	-	-	-
Meso	2	-	3±0	-	-	5
UB	3	-	2	2	-	3
LB	2	24	4±1	7	6	18

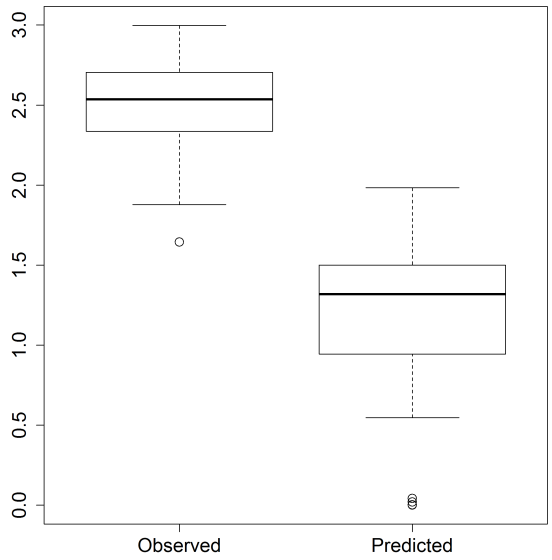

Figure 3.8: Distribution of Shannon diversity of Archaea for both observed and predicted datasets.

3.4 Modelling

The developed model applied to the dataset of 16S rRNA of Archaea (Annex: Table 1) resulted in lower Shannon diversity (Figure 3.8; Mann-Whitney: W= 2585, p-value <0.001), evenness (Figure 3.9; Mann-Whitney: W= 2523, p-value <0.001) and Chao richness indices (Figure 3.10; Mann-Whitney: W= 2465.5, p-value <0.001) as compared to the calculated ones based in measured parameters.

Chao richness showed a similar depth pattern in the model as obtained from the actual data, with lower indices (Multiple comparison test after Kruskal-Wallis; p-value <0.05) for the epipelagic and lower bathypelagic (Figure 3.11; observed difference: 13.08; critical difference: 15.38) than in the upper bathypelagic realm.

Modelled Shannon diversity and evenness performed differently than the observed indexes. Modelled diversity and evenness were significantly lower in the lower bathypelagic than in the epipelagic and upper bathypelagic (Figures 3.18, 3.13; Multiple comparison test after Kruskal-Wallis; p-value <0.05). The modelled Chao richness index showed significant differences between the different regions, however, it could not be pinpointed to specific samples (Figure 3.14; p-value= $\sim$ 0.05; Observed differences: SANT-WTRA=20.31, ARCT-WTRA=19.75;

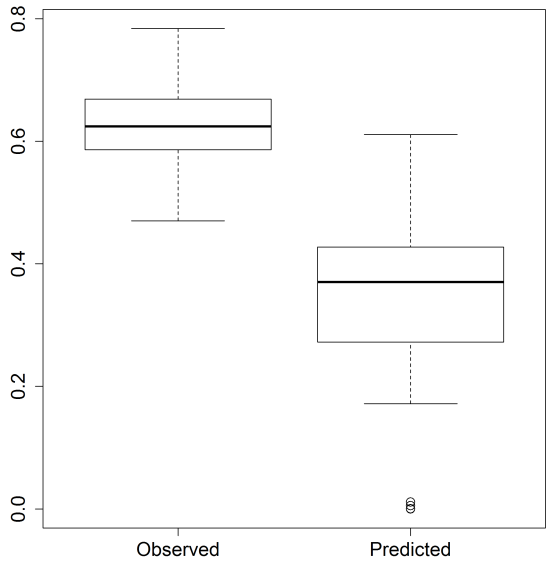

Figure 3.9: Distribution of Shannon evenness of Archaea for both observed and predicted datasets.

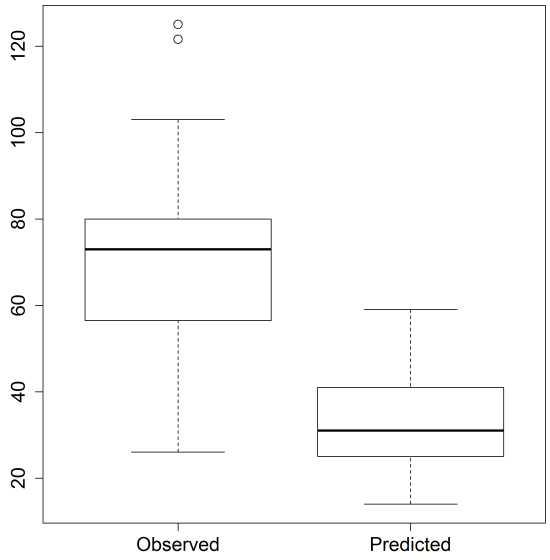

Figure 3.10: Distribution of Chao values of Archaea for both observed and predicted datasets.

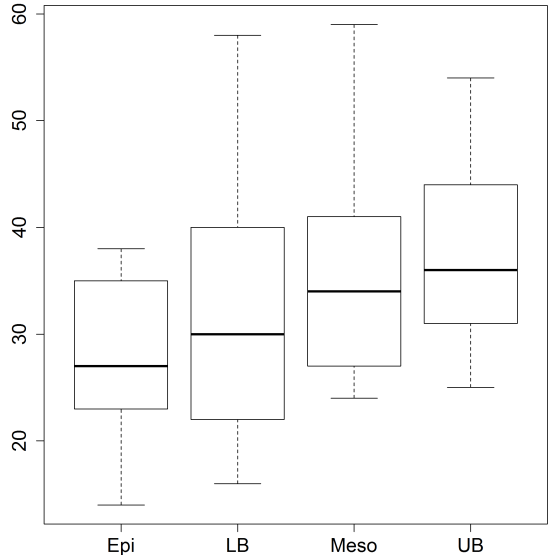

Figure 3.11: Distribution of Chao values of Archaea for the predicted dataset along the sampled depths. Epi - Epipelagic, Meso - Mesopelagic, UB - Upper Bathypelagic, LB - Lower Bathypelagic.

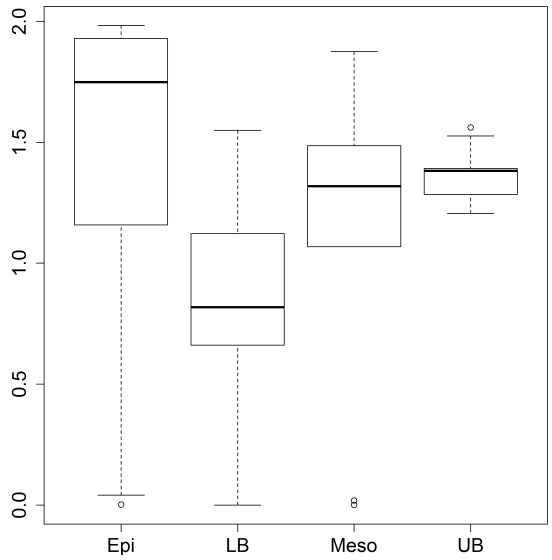

Figure 3.12: Distribution of Shannon diversity of Archaea for the predicted dataset along the sampled depths. Epi - Epipelagic, Meso - Mesopelagic, UB - Upper Bathypelagic, LB - Lower Bathypelagic.

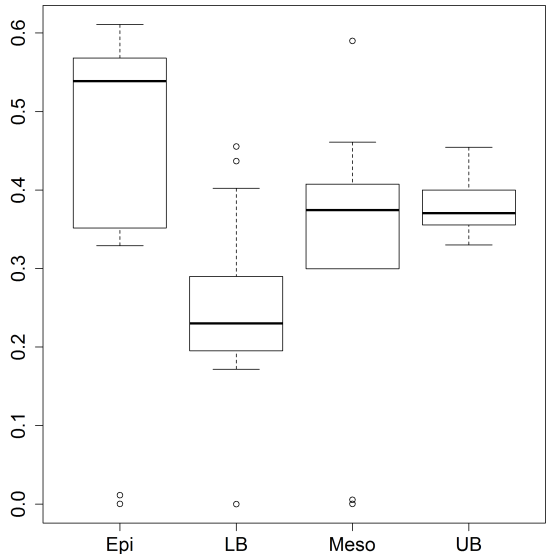

Figure 3.13: Distribution of Shannon evenness of Archaea for the predicted dataset along the sampled depths. Epi - Epipelagic, Meso - Mesopelagic, UB - Upper Bathypelagic, LB - Lower Bathypelagic.

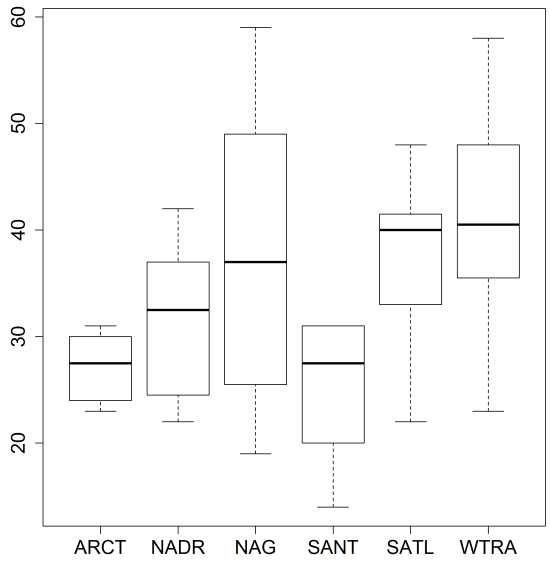

Figure 3.14: Distribution of Chao values of Archaea for the predicted dataset along the sampled provinces. ARCT - North Atlantic Arctic, NADR - North Atlantic Drift, NAG - North Atlantic Gyral, SANT - Subantarctic province, SATL - South Atlantic Gyral, WTRA - Western Tropical Atlantic.

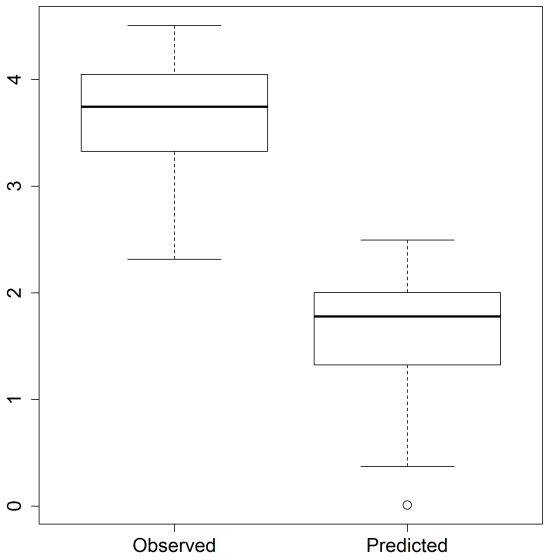

Figure 3.15: Distribution of Shannon diversity of Bacteria for both observed and predicted datasets.

Critical difference= 21.82). The modelled diversity and evenness did not show any significant differences between the different regions.

The model also resulted in lower diversity (Figure 3.15; Mann-Whitney: W= 2598, p-value <0.001), evenness (Figure 3.16; Mann-Whitney: W= 2594, p-value <0.001) and Chao richness indices (Figure 3.17; Mann-Whitney: W= 2548, p-value <0.001) when applied to bacterial *16S rRNA* (Annex: Table 2) as compared to the indices obtained from actual measurements.

Shannon diversity and evenness increased significantly with depth (Multiple comparison test after Kruskal-Wallis; p-value <) with significantly lower values in epipelagic bacterial communities as compared to lower bathypelagic communities (Figure 3.18, 3.19; Observed differences between epipelagic and lower bathypelagic: Shannon diversity=17.44, Shannon evenness=17.81; Critical difference= 15.40). Chao richness, however, did not significantly change with depth. Moreover, no significant difference according to the region was observed for any of the three indices.

The developed model could not be applied to the different *nir*K gene datasets due to the limited dataset.

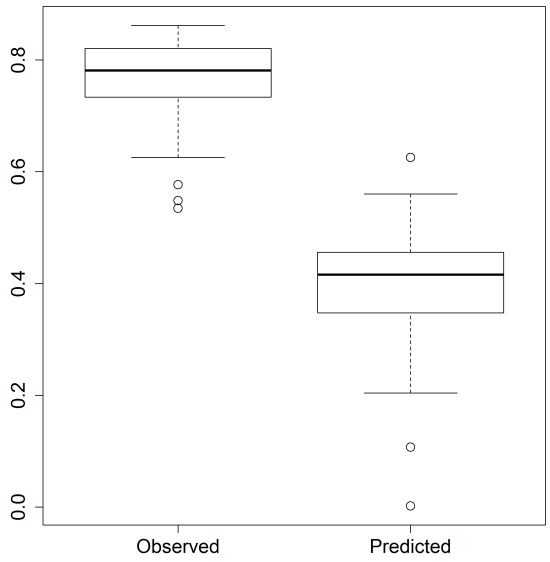

Figure 3.16: Distribution of Shannon evenness of Bacteria for both observed and predicted datasets.

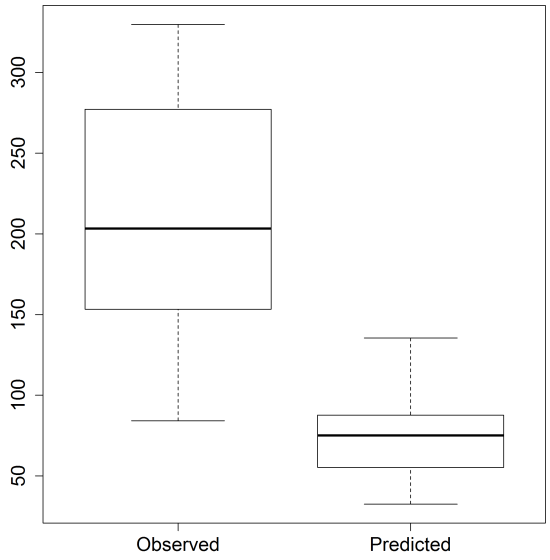

Figure 3.17: Distribution of Chao values of Bacteria for both observed and predicted datasets.

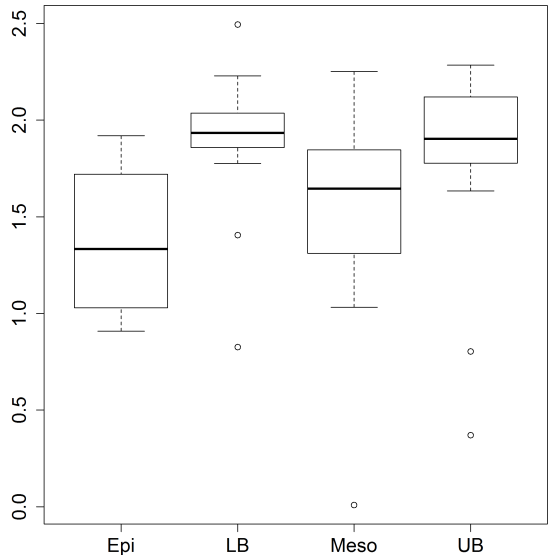

Figure 3.18: Distribution of Shannon diversity of Bacteria for the predicted dataset along the sampled depths. Epi - Epipelagic, Meso - Mesopelagic, UB - Upper Bathypelagic, LB - Lower Bathypelagic.

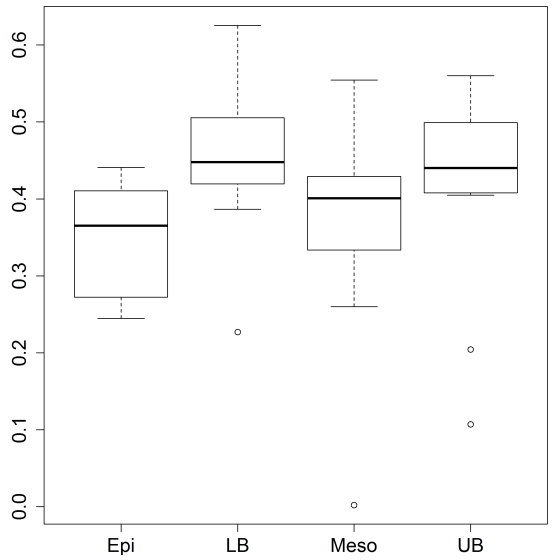

Figure 3.19: Distribution of Shannon evenness of Bacteria for the predicted dataset along the sampled depths. Epi - Epipelagic, Meso - Mesopelagic, UB - Upper Bathypelagic, LB - Lower Bathypelagic.

Chapter 4

Discussion

4.1 Biogeography of prokaryotes.

Biogeography studies the distribution of organisms in space and time. Historically, this scientific discipline has focused on the distribution of macroorganisms, plants and animals. However, in the last decades biogeographic studies of microbes have experienced major advances as a consequence of the development of molecular tools that have allowed the study of uncultured organisms (Olsen *et al.*, 1986; Pace *et al.*, 1986; Woese & Fox, 1977). Several biogeographic patterns described for macroorganisms also occur in prokaryotes, such as latitudinal richness gradients (Fuhrman *et al.*, 2008) and bipolar distribution patterns (Sintes *et al.*, 2015). Moreover, depth distribution patterns of microbes have been frequently described. Not surprisingly, phototrophic (light dependent) prokaryotes inhabit the sunlit layers(Landry & Kirchman, 2002; Partensky *et al.*, 1999) together with microbes benefiting from organic matter released by phytoplankton (Cho *et al.*, 2001), while some other prokaryotic taxa might benefit from the conditions found in the deep ocean (Zeng *et al.*, 2009). Other general biogeographical patterns that prokaryotes exhibit are the latitudinal richness gradients (Fuhrman *et al.*, 2008; Pommier *et al.*, 2007), and the Rapoport rule, *i.e.*, latitudinal ranges of organisms are generally smaller at lower latitudes than at higher latitudes (Amend *et al.*, 2013; Sul *et al.*, 2013).

Our results reveal distinct biogeographical patterns for most bacterial phyla in agreement with previous findings. Chloroflexi (SAR202) was characterized by

higher relative abundances towards low latitudes, while Bacteroidetes, Gammaproteobacteria, Alphaproteobacteria and Marinimicrobia (SAR406) exhibited distribution gradients from one pole to the other (higher relative abundances at one pole and decreasing towards the other), as previously reported for Bacteroidetes and Gammaproteobacteria by fluorescence in situ hybridization in epi- and mesopelagic waters (Schattenhofer *et al.*, 2009).

The pattern observed in Bacteroidetes, Gammaproteobacteria, Alphaproteobacteria and Marinimicrobia (SAR406) might be explained by environmental preferences (*e.g.*, higher concentrations of inorganic nutrients at the northern latitudes), or by the fact that sampling took place in different seasons. Samples collected in the northern hemisphere corresponded to spring and early summer, while samples taken in the southern hemisphere were taken in autumn. However, a previous study determined that the temporal effect on prokaryotic communities was low as compared to spatial and environmental factors in the deep sea (Sintes *et al.*, 2015). The SAR202 metabolism is not well characterized (Schattenhofer *et al.*, 2009) thus it is difficult to explain the distribution of this phylum. However, previous studies (DeLong *et al.*, 2006; Schattenhofer *et al.*, 2009) have reported the increase in relative abundance with depth in Chloroflexi in accordance with our results.

Surprisingly, Nitrospirae and Parcubacteria were relatively abundant in mesopelagic and bathypelagic waters at the southern station of the NAG (North Atlantic Gyre). This region is located close to the Sargasso Sea, characterized by higher primary production than the neighbouring oligotrophic ocean (Jenkins & Goldman, 1985). Members of Nitrospirae are commonly chemolithotrophic nitrite-oxidizing bacteria, while Parcubacteria have been identified mainly in anoxic environments (Nelson & Stegen, 2015). However, little is known about the metabolism of this latter group. Environments favourable for Nitrospirae are associated with high amounts of decaying organic matter in the mesopelagic realm and consequently reduced oxygen concentrations, favouring nitrite oxidizing bacteria. In contrast to the Pacific or Indian Ocean, the oxygen minimum zone in the Atlantic is generally poorly developed in the mesopelagic layer.

Cyanobacteria were found mainly in the euphotic zone where they can perform photosynthesis. However, few sequences associated to this phylum were also found

down to the bathypelagic environment. Thus, the presence of Cyanobacteria sequences in the bathypelagic is most likely attributable to remains of cells sinking down *e.g.* with the faeces of zooplankton (Bruland & Silver, 1981; Caron *et al.*, 1989).

Marinimicrobia (SAR406) with an average abundance of 6 % of the total microbial abundance in the Atlantic increased below the deep chlorophyll maximum in agreement with previous findings (Gordon & Giovannoni, 1996).

The dominance of Proteobacteria has been previously observed (Pham *et al.*, 2008) with Gammaproteobacteria, Alphaproteobacteria and Deltaproteobacteria contributing about 25 %, 24% and 12% of total microbial abundance in marine ecosystems, in agreement with our findings in this study. Deltaproteobacteria, including many sulfur and sulfate-reducing bacteria (Kuever *et al.*, 2005), increased with depth, which might indicate more favourable conditions for sulfur and sulfate-reducing prokaryotes in the deep ocean as compared to the epipelagic realm.

Archaeal taxa did also show biogeographical distribution patterns. Woesearchaeota and Euryarchaeota, the latter in the epipelagic realm and the former in the lower bathypelagic, presented higher relative abundances towards the lower latitudes and in the southern station of NAG and the SATL. The dominance of Euryarchaeota in the epipelagic and a gradual decrease with depth has been previously described (DeLong *et al.*, 2006; Massana *et al.*, 1997, 1998).

Aigarchaeota and marine hydrothermal vent group (MHVG) showed a similar distribution, while Thaumarchaeota displayed an opposite pattern, *i.e.*, higher relative abundances towards the higher latitudes. The presence of MHVG Archaea in specific bathypelagic locations might point to hydrothermal activity close to these sampling sites, as members of this taxon were initially isolated from hydrothermal vents (Takai & Horikoshi, 1999). However, the biogeochemical parameters (Annex 3) and the described local topology of the seafloor in the locations where we detected them, several hundreds to thousands of kilometres away from the Mid Atlantic Ridge or any other active volcanic location (Figures 1.1 and 2.1), do not support a local hydrothermal origin. Thus, the presence of MHVG could indicate the presence of prokaryotes closely related to this group but adapted to different environmental conditions. Alternatively, cells from this

group might originate from a plume located several hundreds of kilometers away and might have been transported to the sampling site, or it could indicate the influence of hydrothermal vents not mapped yet.

The distribution of Thaumarchaeota throughout the Atlantic agrees with previous studies, and suggests that Thaumarchaeota thrives in the oxygen minimum layer (Agogué *et al.*, 2008; Beman *et al.*, 2008) where they could be oxidizing ammonia to support their metabolism (Sintes *et al.*, 2015).

Comparing the communities of Archaea and Bacteria, archaeal community members seem to have a very well defined depth distribution, while for Bacteria although epi- and mesopelagic communities differ from bathypelagic communities, the upper- and lower bathypelagic communities clustered together. Differences in community composition of Archaea were mainly associated to depth, while bacterial community composition was related to oceanic regions. The Arctic bacterial communities differed from the rest of the samples, probably associated to the nutrient-rich environmental conditions (Winter *et al.*, 2013). The bathypelagic layer was inhabited by more similar bacterial communities throughout the ocean, probably due to the more stable and homogeneous environmental conditions (Annex 4). The bacterial communities in the equatorial region (WTRA) were more similar throughout the depth profile, as compared to other regions, probably due to the characteristic upwelling phenomenon in the area, which provides nutrients and Bacteria from deeper layers to the upper layers of the ocean. A previous study (Sintes *et al.*, 2015) suggested that these regions, characterized by deep water mass formation (ARCT and SANT) or upwelling (WTRA), might act as hot spots for dispersion of microbes in the bathypelagic realm, explaining the higher similarity between the communities from different depth layers in these three regions.

4.2 Biogeographical distribution patterns of nitrite reductase harbouring prokaryotes

The *nir*K gene was used as a proxy to characterize the distribution of the group of prokaryotes reducing nitrite. Archaea harbouring nitrite reductase could be

differentiated into two groups according to their *nir*K (Lund *et al.*, 2012), namely *nir*Ka and *nir*Kb. Nitrite reducing Bacteria also express downstream enzymes involved in denitrification processes (Figure 1.3), while the role of *nir*K in Archaea has not yet been completely resolved (Blainey *et al.*, 2011; Hallam *et al.*, 2006; Walker *et al.*, 2010). It is unknown whether nitrite reduction in Archaea is linked to energy conservation (Lund *et al.*, 2012) or nitrite detoxification as described for some Bacteria (Beaumont *et al.*, 2002; Cantera & Stein, 2007).

Both Archaea *nir*K harbouring communities were depth-stratified (Figures 3.5 and 3.6). There was a very strong segregation between the upper bathypelagic, lower bathypelagic and the meso- and epipelagic communities. This stratification could be associated to the higher nutrient supply rates in the epi- and upper mesopelagic as compared to bathypelagic waters, favouring the appearance of archaeal ecotypes as previously described for archaeal ammonia oxidizers (Sintes *et al.*, 2016).

However, archaeal *nir*K a harbouring communities from ARCT and SANT cluster together and do not show depth-stratification, in agreement with the bipolar distribution of Thaumarchaeota (Sintes *et al.*, 2015) and with the function of these locations as hot spot for dispersion of microbes due to deep water mass formation.

In contrast, bacterial *nir*K harbouring communities clustered according to the oceanographic regions (Figure 3.7). Arctic communities from different depth layers clustered together probably due to the deep water mass formation in this region as mentioned above (Annex 4).

4.3 Modelling

Most models that try to explain or predict biogeographical distributions of certain species, they gather information (environmental values) from the observations with (and sometimes without) the species (in our study, OTUs) present as a predictor of presence (or absence, respectively), *i.e.* they compare the values collected from the previous (*e.g.* a subset) with the ones observed on other (new) datasets. If the values on a dataset are close to the samples that had the character present, then the character is most likely present.

These models have two possible outputs, either they state whether a certain character is present or absent (binary output) or give a probability of the presence of the species in a certain point of a dataset. Abundances, however, are not possible to predict with these models.

Here, we tried to develop a model that would predict the abundance of specific microbial OTUs according to environmental conditions. The model developed did not perform well with low abundance OTUs, particularly with OTUs present at only one location and with OTUs that after the rarefaction step were removed. This low performance was due to the fact that the Mahalonabis distance uses the inverse of the covariance matrix of the different environmental parameters, and the calculation does not work with one or fewer points (Mahalanobis, 1936), therefore dominant species were even more dominant, and rarer species became even rarer, which may have been the reason why the overall diversity, evenness and Chao indexes of the predicted values was so low comparing to the observer data. Also, due to the number of different parameters used to calculate the multidimensional statistical space, the calculated Mahalonabis distance resulted in a gross over-estimation of the distance and of the OTU counts in the following steps as they added a lot of noise in the calculation of this distance. This pitfall makes difficult to predict the abundance of OTUs present in extreme environments, *i.e.*, present in only one sample where the maximum or minimum values in a set of parameters occur, such as the samples from higher latitudes or bathypelagic.

Aside from these drawbacks, the patterns in diversity indices calculated from the model were quite similar to the observations. Aspects that could be improved in the model would be including ocean circulation models (Bardin *et al.*, 2014), use of fewer environmental parameters (*e.g.*, temperature, salinity, dissolved oxygen, silicate, phosphate and nitrate) to extrapolate for the dataset of the World Ocean Atlas (Levitus & Technical, 2013).

References

Agogué, H., Brink, M., Dinasquet, J. & Herndl, G.J. (2008). Major gradients in putatively nitrifying and non-nitrifying archaea in the deep north atlantic. *Nature*, **456**, 788–791. 44

Amend, A.S., Oliver, T.A., Amaral-Zettler, L.A., Boetius, A., Fuhrman, J.A., Horner-Devine, M.C., Huse, S.M., Welch, D.B.M., Martiny, A.C., Ramette, A., Zinger, L., Sogin, M.L. & Martiny, J.B.H. (2013). Macroecological patterns of marine bacteria on a global scale. *Journal of Biogeography*, **40**, 800–811. 7, 41

Azam, F. & Worden, A.Z. (2004). Microbes, molecules, and marine ecosystems. *Science*, **303**, 1622–1624. 3

Bardin, A., Primeau, F. & Lindsay, K. (2014). An offline implicit solver for simulating prebomb radiocarbon. *Ocean Modelling*, **73**, 45–58. 46

Beaumont, H.J., Hommes, N.G., Sayavedra-Soto, L.A., Arp, D.J., Arciero, D.M., Hooper, A.B., Westerhoff, H.V. & van Spanning, R.J. (2002). Nitrite reductase of nitrosomonas europaea is not essential for production of gaseous nitrogen oxides and confers tolerance to nitrite. *Journal of bacteriology*, **184**, 2557–2560. 45

Beman, J.M., Popp, B.N. & Francis, C.A. (2008). Molecular and biogeochemical evidence for ammonia oxidation by marine crenarchaeota in the gulf of california. *The ISME Journal*, **2**, 429–441. 44

REFERENCES

Blainey, P.C., Mosier, A.C., Potanina, A., Francis, C.A. & Quake, S.R. (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLOS ONE*, **6**, 1–12. 45

Bruland, K.W. & Silver, M.W. (1981). Sinking rates of fecal pellets from gelatinous zooplankton (salps, pteropods, doliolids). *Marine Biology*, **63**, 295–300. 43

Cantera, J.J.L. & Stein, L.Y. (2007). Role of nitrite reductase in the ammonia-oxidizing pathway of nitrosomonas europaea. *Archives of Microbiology*, **188**, 349–354. 45

Carbonero, F., Oakley, B.B. & Purdy, K.J. (2014). Metabolic flexibility as a major predictor of spatial distribution in microbial communities. *PLoS ONE*, **9**, 1–6. 7

Caron, D.A., Madin, L.P. & Cole, J.J. (1989). Composition and degradation of salp fecal pellets: implications for vertical flux in oceanic environments. *Journal of Marine Research*, **47**, 829–850. 43

Chen, J. (2012). Gunifrac: generalized unifrac distances. *R package version*, **1**, 2012. 18

Cho, B.C., Park, M.G., Shim, J.H. & Choi, D.H. (2001). Sea-surface temperature and f-ratio explain large variability in the ratio of bacterial production to primary production in the yellow sea. *Marine Ecology Progress Series*, **216**, 31–41. 41

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, **311**, 1283–1287. 4

De Corte, D., Sintes, E., Yokokawa, T., Lekunberri, I. & Herndl, G.J. (2016). Large-scale distribution of microbial and viral populations in the south atlantic ocean. *Environmental microbiology reports*. 11

DeLong, E.F., Preston, C.M., Mincer, T., Rich, V., Hallam, S.J., Frigaard, N.U., Martinez, A., Sullivan, M.B., Edwards, R., Brito, B.R. *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science*, **311**, 496–503. 42, 43

Dray, S., Dufour, A.B. *et al.* (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software*, **22**, 1–20. 18

Fierer, N., McCain, C.M., Meir, P., Zimmermann, M., Rapp, J.M., Silman, M.R. & Knight, R. (2011). Microbes do not follow the elevational diversity patterns of plants and animals. *Ecology*, **92**, 797–804. 7

Franklin, M.P., McDonald, I.R., Bourne, D.G., Owens, N.J.P., Goddard, R.C.U. & Murrell, J.C. (2005). Bacterial diversity in the bacterioneuston (sea surface microlayer): the bacterioneuston through the looking glass. *Environmental Microbiology*, **7**, 723–736. 3

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L. & Brown, J.H. (2008). A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 7774–7778. 6, 41

Gardner, M. (1970). Mathematical games the fantastic combinations of john conway's new solitaire game "life". *Scientific American*, **223**, 120–123. 8

Gordon, D. & Giovannoni, S. (1996). Detection of stratified microbial populations related to chlorobium and fibrobacter species in the atlantic and pacific oceans. *Applied and Environmental Microbiology*, **62**, 1171–1177. 43

The geotraces intermediate data product 2014. 9

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y.i., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M. & DeLong, E.F. (2006). Genomic analysis of the uncultivated marine crenarchaeote cenarchaeum symbiosum. *Proceedings of the National Academy of Sciences*, **103**, 18296–18301. 45

Hanage, W.P., Spratt, B.G., Turner, K.M. & Fraser, C. (2006). Modelling bacterial speciation. *Philosophical Transactions of the Royal Society B*, **361**, 2039–2044. 8

Herndl, G.J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A. & Pernthaler, J. (2005). Contribution of archaea to total prokaryotic production in the deep atlantic ocean. *Applied and Environmental Microbiology*, **71**, 2303–2309. 11

Hijmans, R.J. & Van Etten, J. (2014). raster: Geographic data analysis and modeling. r package version 2.2-31. *URL http://CRAN. R-project. org/package= raster.[accessed 15 March 2014]*. 18

Holmes, D.E., Nevin, K.P. & Lovley, D.R. (2004). Comparison of 16s rrna, nifd, reca, gyrb, rpob and fusa genes within the family geobacteraceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, **54**, 1591–1599. 16

Jenkins, W.J. & Goldman, J.C. (1985). Seasonal oxygen cycling and primary production in the sargasso sea. *Journal of Marine Research*, **43**, 465–491. 42

Kirchman, D.L. (2008). *Microbial Ecology of the Oceans*. Wiley-Blackwell, 2nd edn. 3

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. & Glöckner, F.O. (2012). Evaluation of general 16s ribosomal rna gene pcr primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 1–11. 16

Kuever, J., Rainey, F.A. & Widdel, F. (2005). *Class IV. Deltaproteobacteria class nov.*. Springer. 43

Landry, M.R. & Kirchman, D.L. (2002). Microbial community structure and variability in the tropical pacific. *Deep Sea Research Part II: Topical Studies in Oceanography*, **49**, 2669–2693. 41

Lemon, J. (2006). Plotrix: a package in the red light district of r. *R-news*, **6**, 8–12. 18

Levitus, S. & Technical, A.M. (2013). *World Ocean Atlas 2013 (vol:1-4)*. NOAA Atlas NESDIS, 73rd edn. 46

Longhurst, A.R. (2007). *Ecological geography of the sea*. Elsevier Inc. 9, 11

Lund, M.B., Smith, J.M. & Francis, C.A. (2012). Diversity, abundance and expression of nitrite reductase (nirk)-like genes in marine thaumarchaea. *The ISME Journal*, **6**, 1966–1977. 16, 45

Madigan, M.T., Martinko, J.M., Stahl, D.A. & Clark, D.P. (2012). *Brock Biology of Microorganisms*. Benjamin Cummings, 13th edn. 6

Mahalanobis, P.C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, **2**, 49–55. 17, 46

Majaneva, M., Hyytiäinen, K., Varvio, S.L., Nagai, S. & Blomster, J. (2015). Bioinformatic amplicon read processing strategies strongly affect eukaryotic diversity and the taxonomic composition of communities. *PloS one*, **10**, e0130035. 20

Margesina, R. & Miteva, V. (2011). Diversity and ecology of psychrophilic microorganisms. *Research in Microbiology*, **162**, 346–361. 3

Massana, R., Murray, A.E., Preston, C.M. & DeLong, E.F. (1997). Vertical distribution and phylogenetic characterization of marine planktonic archaea in the santa barbara channel. *Applied and environmental microbiology*, **63**, 50–56. 43

Massana, R., Taylor, L.T., Murray, A.E., Wu, K.Y., Jeffrey, W.H. & DeLong, E.F. (1998). Vertical distribution and temporal variation of marine planktonic archaea in the gerlache strait, antarctica, during early spring. *Limnology and Oceanography*, **43**, 607–617. 43

REFERENCES

Mincer, T.J., Church, M.J., Taylor, L.T., Preston, C., Karl, D.M. & DeLong, E.F. (2007). Quantitative distribution of presumptive archaeal and bacterial nitrifiers in monterey bay and the north pacific subtropical gyre. *Environmental Microbiology*, **9**, 1162–1175. 16

Mosier, A.C. & Francis, C.A. (2010). Denitrifier abundance and activity across the san francisco bay estuary. *environmental microbiology reports*, **2**, 667–676. 16

Nakagawa, T., Nakagawa, S., Inagaki, F., Takai, K. & Horikoshi, K. (2004). Phylogenetic diversity of sulfate-reducing prokaryotes in active deep-sea hydrothermal vent chimney structures. *FEMS Microbiology Letters*, **232**, 145–152. 3

Nelson, W.C. & Stegen, J.C. (2015). The reduced genomes of parcubacteria (od1) contain signatures of a symbiotic lifestyle. *Frontiers in Microbiology*, **6**. 42

Newton, I. (1687). *Philosophiæ Naturalis Principia Mathematica*. Royal Society, 1st edn. 8

Olsen, G.J., Lane, D.J., Giovannoni, S.J., Pace, N.R. & Stahl, D.A. (1986). Microbial ecology and evolution: a ribosomal rna approach. *Annual reviews in microbiology*, **40**, 337–365. 41

Pace, N.R., Stahl, D.A., Lane, D.J. & Olsen, G.J. (1986). The analysis of natural microbial populations by ribosomal rna sequences. In *Advances in microbial ecology*, 1–55, Springer. 41

Paradis, E., Claude, J. & Strimmer, K. (2004). Ape: analyses of phylogenetics and evolution in r language. *Bioinformatics*, **20**, 289–290. 18

Parnell, J.J., Rompato, G., Latta, L.C., IV, Pfrender, M.E., Van Nostrand, J.D., He, Z., Zhou, J., Andersen, G., Champine, P., Ganesan, B. & Weimer, B.C. (2010). Functional biogeography as evidence of gene transfer in hypersaline microbial communities. *PLoS ONE*, **5**, 1–8. 7

Partensky, F., Hess, W.R. & Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology and molecular biology reviews*, **63**, 106–127. 41

Pham, V.D., Konstantinidis, K.T., Palden, T. & DeLong, E.F. (2008). Phylogenetic analyses of ribosomal dna-containing bacterioplankton genome fragments from a 4000 m vertical profile in the north pacific subtropical gyre. *Environmental microbiology*, **10**, 2313–2330. 43

Pierce, D. (2012). ncdf4: Interface to unidata netcdf (version 4 or earlier) format data files. *R package, URL http://CRAN. R-project. org/package= ncdf4*. 18

Pommier, T., Canbäck, B., Riemann, L., Boström, K.H., Simu, K., Lundberg, P., Tunlid, A. & Hagström, . (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Molecular Ecology*, **16**, 867–880. 6, 41

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. 18

Rahmstorf, S. (2006). *Thermohaline Ocean Circulation. In: Encyclopedia of Quaternary Sciences*. Elsevier Inc. 3

Rossum, G.v. & Drake Jr., F.L. (2006). *The Python Language Reference Manual*. Python Software Foundation, ISBN 0954161785. 18

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A. & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the atlantic ocean. *Environmental Microbiology*, **11**, 2078–2093. 42

Schippers, A., Neretin, L.N., Kallmeyer, J., Ferdelman, T.G., Cragg, B.A., John Parkes, R. & Jorgensen, B.B. (2005). Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature*, **433**, 861–864. 3

Schliep, K.P. (2011). phangorn: phylogenetic analysis in r. *Bioinformatics*, **27**, 592–593. 18

Schrödinger, E. (1926). An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, **28**, 1049–1070. 8

Sintes, E., De Corte, D., Ouillon, N. & Herndl, G.J. (2015). Macroecological patterns of archaeal ammonia oxidizers in the atlantic ocean. *Molecular Ecology*, **24**, 4931–4942. 7, 9, 41, 42, 44, 45

Sintes, E., De Corte, D., Haberleitner, E. & Herndl, G.J. (2016). Geographic distribution of archaeal ammonia oxidizing ecotypes in the atlantic ocean. *Frontiers in microbiology*, **7**, 77. 45

Stevenson, A. (2015). *Oxford Dictionary of English*. Oxford University Press, 3rd edn. 7

Sul, W., Oliver, T., Ducklow, H., Amaral-Zettler, L. & Sogin, M. (2013). Marine bacteria exhibit a bipolar distribution. *Proceedings of the National Academy of Sciences of the United States of America*, **4**, 2342–2347. 7, 41

Takai, K. & Horikoshi, K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics*, **152**, 1285–1297. 43

Takai, K. & Horikoshi, K. (2000). Rapid detection and quantification of members of the archaeal community by quantitative pcr using fluorogenic probes. *Applied and Environmental Microbiology*, **66**, 5066–5072. 16

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S.W. & Garant, D. (2014). Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology*, **23**, 2890–2901. 7

Walker, C.B., de la Torre, J.R., Klotz, M.G., Urakawa, H., Pinel, N., Arp, D.J., Brochier-Armanet, C., Chain, P.S.G., Chan, P.P., Gollabgir, A., Hemp, J., Hügler, M., Karr, E.A., Könneke,

M., Shin, M., Lawton, T.J., Lowe, T., Martens-Habbena, W., Sayavedra-Soto, L.A., Lang, D., Sievert, S.M., Rosenzweig, A.C., Manning, G. & Stahl, D.A. (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences*, **107**, 8818–8823. 45

Weisburg, W.G., Barns, S.M., Pelletier, D.A. & Lane, D.J. (1991). 16s ribosomal dna amplification for phylogenetic study. *Journal of bacteriology*, **173**, 697–703. 6

Whitman, W.B., Coleman, D.C. & Wiebe, W.J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, **95**, 6578–6583. 3

Winter, C., Matthews, B. & Suttle, C.A. (2013). Effects of environmental variation and spatial distance on bacteria, archaea and viruses in sub-polar and arctic waters. *The ISME journal*, **7**, 1507–1518. 44

Woese, C.R. & Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proceedings of the National Academy of Sciences*, **74**, 5088–5090. 41

Woese, C.R., Kandler, O. & Wheelis, M.L. (1990). Towards a natural system of organisms: Proposal for the domains archaea, bacteria, and eucarya. *Proceedings of the National Academy of Sciences of the United States of America*, **87**, 4576–4579. 3

Zeng, X., Birrien, J.L., Fouquet, Y., Cherkashov, G., Jebbar, M., Querellou, J., Oger, P., Cambon-Bonavita, M.A., Xiao, X. & Prieur, D. (2009). Pyrococcus ch1, an obligate piezophilic hyperthermophile: extending the upper pressure-temperature limits for life. *The ISME journal*, **3**, 873–876. 41

Zwietering, M.H., Jongenburger, I., Rombouts, F.M. & van 't Riet, K. (1990). Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, **56**, 1875–1881. 8

Annex

Listing 1: Model script

```
#########################################################################################
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++#
#+++++++++++++++++Built by Miguel Fernandes Guerreiro+++++++++++++++++++++++++++++++++#
#+++++++++++++++++++++++++++++++++27/09/2016+++++++++++++++++++++++++++++++++++++++++#
#+++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++#
#########################################################################################
#modules needed
from matrix_Builder import file2Matrix
import numpy as np
##from numpy import cov,array
from scipy.spatial.distance import mahalanobis
import sys

#functions
def MatrixPonderation(Counts,Env,OTUnum):
    """Produces a matrix with environmental
values with a number of rows equal to the
number of counts found for the species.
Requires:Counts is a list of lists with counts
 of otus per row/first level list) vs sampling
  site(per column/second level list) and
  Environmental Table is a list
of lists with Environment parameters per
row/first level list) vs sampling site (per
 column/second level list) and OTU row in
  Count table(int).
Ensures:Table with number of rows(samples)
 equal to its counts(list of lists).
```

```
"""
    from matrix_Builder import file2Matrix

    PonderatedMatrix=[]
    row=[]
    count=0
    for r in range(0,len(Env)):
        count=int(Counts[OTUnum][r])
        if count!=0:
            for c in range(0,len(Env[0])):
                row.append(float(Env[r][c][1:-1]))
                #data has '"float"' string
            while count>0:
                PonderatedMatrix.append(row)
                count-=1
            row=[]
    return PonderatedMatrix


def StandardVectors(neigh):
    """matrix mean of values
Requires:
Ensures:Returns averages(list)
"""
    import copy
    neighboorhood=copy.deepcopy(neigh)
    listm=[]
    mean=0
    ###starting the overarching operation
    for c in range(0,len(neighboorhood[0]),1):
        for r in range(0,len(neighboorhood),1):
            mean+=float(neighboorhood[r][c])#sum for mean
        ###mean calculation and list
        try:
            mean=mean/float(len(neighboorhood))
        except ZeroDivisionError:
            if mean!=0:
                raise Exception('something_wrong')
        listm.append(mean)

        ###preparetion for next turn
```

```
        mean=0
    del mean

    return listm


def CountsDistrib(Counts,OTUsnum):
    """receives Counts files and row of OTU,
    and retrieves the average and SD
Requires:Counts is a list of lists with counts
of otus per row/first level list)
vs sampling site(per column/second level
list), OTUsnum(int)
Ensures:tuple with mean(float) and
sd(float)."""
    listsd=[]
    otusMean=0
    #data has string of taxonomy in last element of row list
    for c in Counts[OTUsnum][0:-1]:
        listsd.append(int(c))
        otusMean+=int(c)
    otusMean=otusMean/float(len(Counts[0]))
    otusSD=0
    for i in listsd:
        otusSD+=(i-otusMean)**2
    otusSD/=len(listsd)
    return otusMean,otusSD


def calculateCount(Counts,Mean,SD,Dist,OTUsnum,SampleN):
    """
Requires:Counts is a list of lists
(with counts of otus per row/first level list)
vs sampling site(per column/second level list)
,Mean(Float) of counts of OTU,
SD(Float) of counts of OTU,Dist(Float)
distance of sample site SampleN to average
sample site, OTUsnum(int) row number of
 the OTU in Counts list of lists,
SampleN(Int) Column number of the
Sample site in Counts list of lists.
```

```
Ensures:Count calculated through
Baas-Becking principle 'Everything is
everywhere, the environment selects'.
"""
    Count=int(Counts[OTUsnum][SampleN])
    count1=Mean+(SD*Dist)
    count2=Mean-(SD*Dist)
    cc=0
    if abs(Count-count1)<abs(Count-count2):
        cc=count1
    elif abs(Count-count1)>abs(Count-count2):
        cc=count2
    if cc<0:
        cc=0
    return int(round(cc))


#Program body


Countsfile='NBOf.txt'#raw_input('Counts file name')
Envfile='envdataseqf.txt'#raw_input('env file name')
Counts=file2Matrix(Countsfile, Cseparator='\t',
 Keepheader=False,
 KeepFirstColumn=False)#tax column
Env=file2Matrix(Envfile, Cseparator='\t',
 Keepheader=True,
 KeepFirstColumn=False)#header and 1st column check
nOTUS=len(Counts)
nSAMPLES=len(Counts[0])
table=''
poin=[]
otusOUT=0
otusOU=0
for r in range(0,nOTUS):
    neigh=MatrixPonderation(Counts,Env,r)
    try:
        m=np.array(StandardVectors(neigh))#V
    except IndexError:
        otusOUT+=1
        print r#no observations
```

```
            continue
        neigh=np.array(neigh)
        try:
            VI=np.linalg.inv(np.cov(neigh, rowvar=False).T)#meh
        except:
            otusOU+=1
            print '\t',r#observation in 1 place
            continue
        for s in range(0,nSAMPLES-1):
            for a in Env[s]:
                poin.append(float(a[1:-1]))#data has ' "float" ' string

            poin=np.array(poin)
            Dist=mahalanobis(poin,m,VI)#
            poin=[]
            Mean,SD=CountsDistrib(Counts,r)

            table+=str(calculateCount(Counts,Mean,SD,Dist,r,s))+'\t'
        table=table[:-1]+'\n'

print otusOUT,otusOU
handle=open('NBObb.txt','w')
handle.write(table)
handle.close()
```

Table 1: Model on 16S Archaea results. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	2.38±0.54	3.0±0.04	1.7±0.05	3.05±0.22	2.59±0.26	2.73±0.83
Meso	3.46±0.26	2.59±1.12	3.64±1.03	2.17±0.01	2.79±0.09	2.41±0.59
UB	1.35±0.15	5.22±0.68	4.01±0.18	3.09±0.27	3.99±0.66	3.36±0.21
LB	3.1±1.21	3.55±0.05	3.41±0.45	3.77±0.27	3.9±0.07	4.22±0.16

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.04±0.0	0.05±0.01	0.04±0.0	0.06±0.01	0.07±0.01	0.04±0.01
Meso	0.06±0.01	0.04±0.02	0.06±0.02	0.04±0.0	0.05±0.0	0.04±0.01
UB	0.03±0.0	0.08±0.02	0.05±0.0	0.03±0.01	0.05±0.01	0.06±0.01
LB	0.06±0.02	0.05±0.0	0.06±0.01	0.05±0.0	0.05±0.0	0.06±0.0

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	4±1	6±1	3±0	5±1	4±0	5±1
Meso	7±1	5±2	6±1	4±0	4±1	4±1
UB	2±0	10±0	8±1	6±0	8±1	6±0
LB	7±3	6±0	6±1	7±0	7±0	9±0

Table 2: Model on 16S Bacteria results. A) Shannon Diversity, B) Shannon Eveness and C) Chao. Epi - epipelagic, meso - Mesopelagic, UB - upper bathipelagic, LB - lower bathipelagic.

A)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	2.38±0.54	3.0±0.04	1.7±0.05	3.05±0.22	2.59±0.26	2.73±0.83
Meso	3.46±0.26	2.59±1.12	3.64±1.03	2.17±0.01	2.79±0.09	2.41±0.59
UB	1.35±0.15	5.22±0.68	4.01±0.18	3.09±0.27	3.99±0.66	3.36±0.21
LB	3.1±1.21	3.55±0.05	3.41±0.45	3.77±0.27	3.9±0.07	4.22±0.16

B)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	0.04±0.0	0.05±0.01	0.04±0.0	0.06±0.01	0.07±0.01	0.04±0.01
Meso	0.06±0.01	0.04±0.02	0.06±0.02	0.04±0.0	0.05±0.0	0.04±0.01
UB	0.03±0.0	0.08±0.02	0.05±0.0	0.03±0.01	0.05±0.01	0.06±0.01
LB	0.06±0.02	0.05±0.0	0.06±0.01	0.05±0.0	0.05±0.0	0.06±0.0

C)

Layer	ARCT	NADR	NAG	WTRA	SATL	SANT
Epi	4±1	6±1	3±0	5±1	4±0	5±1
Meso	7±1	5±2	6±1	4±0	4±1	4±1
UB	2±0	10±0	8±1	6±0	8±1	6±0
LB	7±3	6±0	6±1	7±0	7±0	9±0

Table 3: Tables with average and standard deviation of environmental parameter: A) entire transect and B) lower bathypelagic layer.

A)

Environmental Parameter	mean±sd
TALK(μmol/kg)	2333.39±29.05
Al(nmol/kg)	11.74±9.42
Cd(nmol/kg)	0.27±0.22
Fe(nmol/kg)	0.51±0.31
Mn(nmol/kg)	0.44±0.54
Ni(nmol/kg)	4.06±1.46
Pb(pmol/kg)	18.83±9.57
Zn(nmol/kg)	1.62±1.90
Y(pmol/kg)	138.86±26.75
La(pmol/kg)	23.12±9.31
Salinity	35.17±0.62
Oxygen(μmol/kg)	216.96±38.43
Fluorescence(arb)	0.05±0.08
^{3}H-Leucine uptake(pmol/L/d)	48.66±91.90
PO_4^{3-}(μmol/kg)	1.08±0.68
Si(μmol/kg)	21.29±29.74
NO_2^-(μmol/kg)	0.05±0.13
NO_3^-(μmol/kg)	16.11±10.22

B)

Environmental Parameter	mean±sd
TALK(μmol/kg)	2335.63±19.88
Al(nmol/kg)	16.79±9.20
Cd(nmol/kg)	0.41±0.20
Fe(nmol/kg)	0.55±0.09
Mn(nmol/kg)	0.18±0.10
Ni(nmol/kg)	5.07±1.36
Pb(pmol/kg)	9.70±3.42
Zn(nmol/kg)	3.37±2.27
Y(pmol/kg)	167.50±26.09
La(pmol/kg)	34.79±9.22
Salinity	34.84±0.09
Oxygen(μmol/kg)	244.27±15.29
Fluorescence(arb)	0.01±0.01
^{3}H-Leucine uptake(pmol/L/d)	0.93±1.72
PO_4^{3-}(μmol/kg)	1.47±0.47
Si(μmol/kg)	50.17±41.31
NO_2^-(μmol/kg)	0.02±0.01
NO_3^-(μmol/kg)	21.75±6.61

Table 4: Tables with average and standard deviation of environmental parameter: A) entire transect and B) all of the bathypelagic layer .

A)

Environmental Parameter	mean±sd
TALK(μmol/kg)	2333.39±29.05
Al(nmol/kg)	11.74±9.42
Cd(nmol/kg)	0.27±0.22
Fe(nmol/kg)	0.51±0.31
Mn(nmol/kg)	0.44±0.54
Ni(nmol/kg)	4.06±1.46
Pb(pmol/kg)	18.83±9.57
Zn(nmol/kg)	1.62±1.90
Y(pmol/kg)	138.86±26.75
La(pmol/kg)	23.12±9.31
Salinity	35.17±0.62
Oxygen(μmol/kg)	216.96±38.43
Fluorescence(arb)	0.05±0.08
^{3}H-Leucine uptake(pmol/L/d)	48.66±91.90
PO_4^{3-}(μmol/kg)	1.08±0.68
Si(μmol/kg)	21.29±29.74
NO_2^-(μmol/kg)	0.05±0.13
NO_3^-(μmol/kg)	16.11±10.22

B)

Environmental Parameter	mean±sd
TALK(μmol/kg)	2326.12±18.11
Al(nmol/kg)	12.72±8.20
Cd(nmol/kg)	0.41±0.20
Fe(nmol/kg)	0.62±0.15
Mn(nmol/kg)	0.17±0.08
Ni(nmol/kg)	4.98±1.30
Pb(pmol/kg)	17.28±11.78
Zn(nmol/kg)	2.87±1.94
Y(pmol/kg)	154.64±22.78
La(pmol/kg)	28.51±9.12
Salinity	34.84±0.16
Oxygen(μmol/kg)	228.46±25.39
Fluorescence(arb)	0.01±0.01
^{3}H-Leucine uptake(pmol/L/d)	1.13±1.86
PO_4^{3-}(μmol/kg)	1.51±0.47
Si(μmol/kg)	37.92±34.19
NO_2^-(μmol/kg)	0.01±0.01
NO_3^-(μmol/kg)	22.52±6.64